DOCUMENT RESUME

ED 462 735 EA 031 547

AUTHOR	Mitchell, Douglas E.; Mitchell, Ross E.
TITLE	Competing Explanations of Class Size Reduction Effects: The California Case.
PUB DATE	2001-04-14
NOTE	61p.; Paper presented at the Annual Meeting of the American Educational Research Association (Seattle, WA, April 10-14, 2001).
PUB TYPE	Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	Academic Achievement; *Class Size; Classroom Environment; Classroom Research; Elementary Secondary Education; *Small Classes; Socialization
IDENTIFIERS	*California

ABSTRACT

Competing explanations of class size reduction effects on student academic achievement were tested using student, teacher, and school data collected from nearly 700 classrooms in over 70 schools during the first 3 years of implementation of California's (K-3) Class Size Reduction Program. Five major hypotheses were tested: (1) overall impact of class-size reduction is greater when exposure is longer; (2) academic socialization of students is greater when reduced-size class experiences begin in the earliest grades (K-1); (3) reduced classroom management overhead in smaller classes leads to higher performance; (4) school instructional resource utilization is more effective at raising achievement in smaller classes; and (5) changes in instructional practice result in changing the pattern of student achievement outcomes in small classes such that class performance is more uniform as well as higher overall. The California experience suggests that longer and earlier class-size reduction provides modest achievement benefits, but there are no differentially greater benefits for at-risk/disadvantaged students. School resource utilization does not appear to be more effective. Classroom teachers' practices appear to be moving the bulk of the middle of the class toward the higher performing students in the achievement distribution, but only slightly. Appendices describe variables and models used in the study. (Contains 61 references.) (Author/RT)

Competing Explanations of Class Size Reduction Effects:

The California Case

Douglas E. Mitchell and Ross E. Mitchell

Graduate School of Education

University of California

Riverside, California 92521

Paper presented at the Annual Meeting of the

American Educational Research Association

Seattle, Washington

April 14, 2001

ABSTRACT

Competing explanations of class size reduction effects on student academic achievement are tested using student, teacher, and school data collected from nearly 700 classrooms in over 70 schools in seven districts during the first three years of implementation of California's (K-3) *Class Size Reduction Program.* Five major hypotheses are tested by examining the multi-level dependence of third and fourth grade students' performance in total mathematics on the Stanford Achievement Test (9^{th} Edition – Form T) at the end of the third year of implementation: 1) the overall impact of class size reduction is greater when exposure is longer; 2) the academic socialization of students is greater when reduced size class experiences begin in the earliest grades (K/1); 3) reduced classroom management overhead in smaller classes leads to higher performance; 4) school instructional resource utilization is more effective at raising achievement in smaller classes; and 5) instructional practice changes result in changing the pattern of student achievement outcomes in small classes such that class performance is more uniform as well as higher overall. The California experience suggests that longer and earlier class size reduction experiences provide modest achievement benefits, but there are no differentially greater benefits for at-risk/disadvantaged students as a consequence of prolonged exposure, early socialization, or reduced classroom management overhead. School resource utilization does not appear to be more effective. Classroom teacher practices appear to be moving the bulk of the middle of the class toward the higher performing students in the achievement distribution, but only slightly.

In California, the *Class Size Reduction Program* authorized by Senate Bill 1777 in 1996 continues to represent the most expensive educational reform effort ever undertaken by any state, reducing class size from an average of 29 to 19 for 92% of the eligible students in kindergarten through grade three by the third year of implementation. State funds allocated during the first three years of operation amounted to nearly $4.1 billion – about $3.3 billion for operation, with an additional $0.8 billion required for school facilities (Table 1). These figures do not include any expenditures from local school district general funds that may have been needed to offset excess staff or facilities costs.

Table 1. State Funding Allocations by Category for the *California Class Size Reduction Program,* grades K-3, from 1996-1997 through 1998-1999

School Year	Operations	Facilities	Total	Cumulative Total
1996-1997	$611,275,000	$342,802,500	$954,077,500	$954,077,500
1997-1998	$1,216,587,000	$311,628,438	$1,528,215,438	$2,482,292,938
1998-1999	$1,439,456,096	$154,360,000	$1,593,816,096	$4,076,109,034

Sources: The following documents were downloaded on January 22, 2001 from the California Department of Education Class Size Reduction website:
http://www.cde.ca.gov/classsize/particip/sum96.htm,
http://www.cde.ca.gov/classsize/particip/sum97.htm,
http://www.cde.ca.gov/classsize/particip/sum98.htm, and
http://www.cde.ca.gov/classsize/facts.htm

This report documents ongoing evaluation of the impact of the California class size reduction initiative sponsored by the California Educational Research Cooperative (for earlier CERC studies, see D. Mitchell & R. Mitchell, 1999, 2001; R. Mitchell, 2000; Ogawa, Huston, & Stine, 1999). Currently, we are investigating how class size is linked to student academic achievement outcomes. Initial findings from the evaluation of student standardized test

performance measured in the second year of implementation indicated that effects were uneven, and the negative consequences of disruption created by rapidly reducing class sizes in the first year were at least as great as any benefits that might have accrued (D. Mitchell & R. Mitchell, 1999). Reduced class size experiences in the second year appeared to offer a positive benefit, but this improvement was rather small compared to the negative consequences associated with coming from a poor or non-English speaking home, being a member of an "under-represented" racial/ethnic group, and attending a school that has a significant number of teachers without full certification (R. Mitchell, 2000). In the previous CERC study, the second year of implementation of class size reduction was found to be potent enough to offset the negative achievement consequences of combination grade classrooms.

We will not undertake a lengthy review of the relationship between class size and school outcomes and processes here. Extensive discussions of class size and the outcomes of schooling may be found in treatises by Achilles (1999), Hanushek (1998), and the earlier work by Glass, et al. (1982). A brief summary of the most current discussion of class size impacts on student achievement is provided below.

Recent studies offer important insights into the overall impact of class size on teacher behavior and student achievement. Tennessee *Project STAR* data continue to be reanalyzed, including various efforts to follow the 1985 cohort through later elementary, middle, and high school years (e.g., Blatchford, Goldstein, & Mortimore, 1998; Finn, Gerber, Achilles, & Boyd-Zaharias, in press; Goldstein & Blatchford, 1997; Hanushek, 1999; Krueger, 1999; Krueger & Whitmore, 2001; Nye, Hedges, & Konstantopoulos, 1999, 2000; Pate-Bain, et al. 1997). These improvements and refinements reconfirm earlier analyses indicating that Tennessee's class size reduction (CSR) experiment was successful at facilitating improved student performance. But

there is some reason to believe that the effects of CSR in Tennessee were slightly less powerful than originally reported. Additionally, the benefits of a small class experience for students who were not enrolled in the program until second or third grade are noticeably less than that obtained by those who started in kindergarten or first grade.. Unfortunately, the single cohort design does not permit a clear distinction between the effects of student mobility and timing of CSR experience because too few students were permitted to violate the design by moving from a large class to a small class while remaining in the same school. Thus, with exhaustive reanalyses, the basic conclusions offered from *Project STAR* remain the same:

- earlier is better (K or first grade),
- longer is better (K/1 through third – at least three years – offers the greatest benefit),
- a more conducive classroom learning environment is produced, and
- positive student achievement, behavior and attitude effects persist, but weaken as students continue through school.

Other recent efforts worthy of attention include the Wisconsin *SAGE* evaluations and a study in England examining class size and the adult-pupil ratio. The Wisconsin program has substantially reproduced the basic outlines of the Tennessee studies: improvement in the first year with the improved performance remaining stable in subsequent years for students enrolled in a class with a reduced student to certified teacher ratio of 15 to 1 – this includes classes with two teachers and 30 students – and a greater benefit to African American students (Molnar, et al., 2000). These results are most notable for mathematics achievement, while benefits in reading and language are smaller. In an examination of the first three years of reporting on *SAGE,* Hruz (2000) cautions that the positive results may be due almost entirely to the benefit to African American students – since white students are not benefiting greatly if at all.

The Wisconsin evaluators are making some effort to attend to teacher disposition and work performance, but their study design does not permit them to make causal inferences about

the link between teacher attitudes and behaviors and student outcomes as a function of class size. A point related to teaching that has not received much attention is that the classes identified as high performing have much higher average teacher experience than the low performing classes (Hruz, 2000). Thus, the question of whether it is the benefit of having experienced teachers or a reduced size class that is more strongly related to student achievement remains open.

A British study also confirms that small classes at the start of school are beneficial to students, and that initially low achieving students benefit most from the experience (Blatchford, 2000). Further, teacher ability and effort to attend to individual pupil needs and performance is increased in a reduced size class, where student attention is better maintained, and disruptive and off-task behavior is reduced. But an important cautionary note is offered in this study as well. The smaller class size creates a social environment that can lead to more aggressive children or to children being rejected by their peers. Either due to lack of alternative peers, or lack of a perceived need to interact with and learn from peers, the young English children in this study displayed more social adjustment difficulties. Thus, the story is fairly consistent outside of Tennessee, both within and outside of the United States. A small class experience is most effective when students begin school (K/1), most valuable to students who are academically at-risk, and the benefits are more likely to persist if students are in smaller classes longer. But despite the average gains associated with class size reduction, not all small classes are beneficial nor are all large classes detrimental.

In sum, research to date supports seven broad conclusions:

1. The overall effects of CSR are modest in size, and in danger of being obscured by other factors influencing student achievement
2. Earlier exposure to CSR is more likely to produce significant achievement gains.
3. Longer participation in small classes does not necessarily produce greater achievement gains, but may make the gains more resistant to decay.

4. The effects of small class experience persist after children return to larger classes, but these effects decay over time.
5. Some populations of students seem to gain more from participation in small classes than others. Specifically, at-risk poor and under-represented minority children seem to show larger gains for the same amount of exposure.
6. While classroom processes and curriculum content are certainly important factors in achievement, it is hard to document specific changes in curriculum and instruction that both accompany reductions in class size and are responsible for achievement gains.
7. The 1999 finding by CERC researchers that California's *Class Size Reduction Program* produced vanishingly small impacts on student achievement as measured by the mandated Stanford Achievement Test – 9th Edition was confirmed by a substantially funded statewide CSR evaluation consortium (Bohrnstedt & Stecher, 1999).

Theoretical Hypotheses

Before detailing our four major theoretical propositions about how class size reduction impacts student achievement, two preliminary considerations are necessary. First, this is a theory-based policy evaluation and not an experimental study. As such, it is necessary to examine the possibility that class size reduction implementation was biased. Biases in implementation are likely to be associated with student achievement. Modeling of student achievement outcomes due to class size reduction must provide an accounting of possible associational biases in order to isolate CSR effects.

H1: ***The Implementation Bias Hypothesis*** – If CSR is implemented as a program rather than as an experiment, there will be significant opportunity biases that have to be controlled before achievement effects can be documented

The second preliminary consideration is the duration of exposure question: Does it matter how long a student receives instruction in a reduced size class? This question has received great attention recently, as reviewed above, and it is necessary to determine if the positive benefits of prolonged exposure to a reduced size class are reproduced in the California experience.

H2: ***The Overall Impact Hypothesis*** – If the benefits of exposure to CSR are uniformly distributed and accruing over time, there will be a dose-response pattern, with longer exposure leading to greater benefits across all subjects.

Now we may proceed with the presentation of our four research hypotheses. Several explanations have been offered for how smaller classes might produce higher academic achievement (for reviews, see Achilles, 1999; Anderson, 1999). These explanations can be classified as emphasizing changes in classroom socialization (Finn 1998; Hruz, 2000; Krueger, 1999), instructional practices (Johnston, 1989; Zahorik 1999), classroom time management (Bain, Lintz, & Word, 1989; Glass, Cahen, Smith, & Filby, 1982; Johnston, 1989), or resource availability (Johnston, 1989; Mitchell, Beach, & Badarak, 1989). These four explanations predict that different *patterns* of student achievement improvement will result as well as increased *average* student attainment.

H3: ***The Socialization Hypothesis*** **-** If the benefits of CSR are produced mainly through better socialization to school, the greatest advantages will go to children with early exposure and to children with the greatest need for socialization to school norms (poor and underrepresented minority students)

The capacity for the teacher to notice and attend to individual students and sustain attention is offered as the mechanism by which students become positively socialized to classroom life and academic expectations. Students become engaged in the business of school with greater success and more positive affect. A smaller class also creates working conditions that lead to reduced stress and greater motivation for teachers, leading to increased task related interactions and fewer routine management interactions, thereby gaining a greater sense of efficacy (Hargreaves, Galton, & Pell, 1997). If classroom socialization is more effective, then students in their first year of school should receive the greatest benefit from a small class with substantially diminishing returns for each successive year. This should be reflected in higher mean achievement in the earliest grades, but should not necessarily have persisting effects in later grades. Achievement should be more markedly improved among students from typically

educationally disadvantaged groups as well, suggesting either a narrowing of the achievement dispersion or at least a more positive skewing (clipping the low performance tail).

> H4: ***The Classroom Management Hypothesis*** –If the benefits of small class exposure are mediated by reductions in classroom management overhead, greater benefits will accrue to classes with challenging management problems and will be reflected in marginal achievement gains (after controlling for prior achievement)

Having to manage fewer disruptions, i.e., less interruption or slowing of classroom routines, is most frequently offered as the source of more time to use for learning activities. If more effective or efficient use of time is responsible for higher achievement, then students should experience a fuller or more extensive curriculum in a smaller size class each year. This should raise the mean performance without necessarily impacting the distribution of achievement. These benefits should be available from initial implementation and not exhibit a significant cohort effect.

> H5: ***The Resource Effectiveness Hypothesis*** **-** If the impact of CSR is mediated through more effective use of resources then current CSR exposure will yield increased marginal gains in the most impoverished schools and among the most challenged students.

The availability of better teachers, more instructional materials, competent peers, and other educational resources will enhance learning. Small classes may attract or retain better teachers, allow greater use of the same class set of materials previously used by a larger class, and reduce the dispersion of student ability in the classroom, thereby creating a more homogeneous group. If resource availability is important, then the benefits of small classes should be greatest for those who attend the most resource poor schools (poor students, poor in facilities and materials, poor teachers, etc.), with possibly some small benefit to smaller classes in resource rich schools as well (more teachers per student, regardless). Resource improvements are likely to make a difference at school entry – getting started right – and have continuing

benefits as well – no lost opportunities. There may be diminishing returns of improved resource availability, but far less dramatic than that for socialization. Resource improvements should be reflected primarily in changes in mean performance.

H6: ***The Instructional Practice Hypothesis*** – If the benefits of CSR are created by providing an opportunity for better use of classroom resources then instruction will be more uniformly effective, leading to reduced dispersion (i.e., lower standard deviation, higher kurtosis and less positive skewing) after controlling for intake achievement patterns.

Individualizing instruction, i.e., accurately meeting student learning needs, is the most popularly cited change in instructional practice. Smaller student groups (Hallinan & Sørensen, 1985) and increases in "hands-on" learning activities (Molnar, Zahorik, & Smith, 1999) are also supposed to be found in smaller classes. If improved instructional practice is responsible for higher achievement, then more students should learn more fully the content of the curriculum in each year of small class experience than they would in a large class. Assuming that high performing students receive a smaller marginal benefit than lower performing students from improved instruction, the range of performance should narrow while the mean increases over successive years. Field research indicates that there is little immediate change in teacher behavior in response to a small class (Borhnstedt & Stecher, 1999; Cahen, Filby, McCutcheon, & Kyle, 1983). Institutional practice effects should be more pronounced as smaller classes have been in place longer. Thus, successive cohorts should reap larger benefits as improved instructional practices become institutionalized.

Cautionary Notes: Accurate Assessment of CSR Impact is Quite Challenging

Five problems are encountered whenever we try to evaluate broad policies like CSR. First, CSR is accompanied by a host of other efforts to improve achievement – the impacts of many of these efforts cannot be easily separated from the impact of changing class size.

California enacted more than a dozen school reform and improvement policies during the same period as the development and implementation of CSR, including:

1) Passage of California Proposition 227 which has sharply curtailed bilingual education programs,
2) Adoption of a statewide accountability policy forcing multiple assessments of student achievement and requiring reports on all students not reaching grade-level achievement standards,
3) Implementation of a Beginning Teacher Support and Assessment program creating a two year induction program for new teachers,
4) Changes in the funding model for special education which substantially affects local district costs when children are certified for services,
5) Changing economic conditions that affect unemployment and poverty rates in many districts,
6) Continued immigration and relocation which changes the composition of many school populations,
7) A broad reading initiative aimed at changing the focus and effectiveness of early literacy instruction,
8) Changes in regulations regarding the certification of teachers that have changed both the character and timing of pre-service teacher preparation,
9) Support for development of new instructional technologies aimed at providing students with better access to location-independent and multi-media learning opportunities,
10) Adoption of a new statewide standardized achievement test (the Stanford Achievement Test, version 9) and mandated school level public reporting of achievement test scores,
11) Continued implementation of new textbook and curriculum materials adoption cycles (both language arts and mathematics curriculum frameworks were changed at the time of CSR policy adoption and implementation) assuring major changes in the scope, sequence and content of subject matter curricula,
12) Addition of ninth grade class size reduction for specific subjects,
13) Changes in regulations regarding the certification of school administrators that have changed both the character and timing of pre-service administrator preparation.
14) Establishment of a powerful Peer Assisted Review (PAR) program aimed at holding experienced teachers accountable for self-improvement.

Second, the impact of reducing class size is entangled with and embedded in a wide range of student demographic, classroom, school and district factors that have powerful effects on achievement making it impossible to make simple direct measurements of the specific contributions of CSR. As a result, statistical analysis has to be used to disentangle the several

contributions to student achievement – but even the best statistical techniques do not give foolproof tests.

Among the most prominent demographic factors that are known to have effects large enough to obscure class size effects are: family poverty, ethnicity, home language, inter-school transiency and student gender (e.g., Entwisle & Alexander, 1992; Han & Hoover, 1994; Jencks & Phillips, 1998; D. Mitchell & R. Mitchell, 1999; Rosenthal, Baker & Ginsburg, 1983). Within schools and classrooms, such factors as grade to grade cohort achievement variations, special education placements, language proficiency levels, combination grade class assignments, and grade-level retention can be expected to influence measured achievement (e.g., Balow & Schwager, 1990; Burns, 1996; Entwisle, Alexander, & Olson, 1997; Hakuta, Butler, & Witt, 2000; Mitchell, Destino, & Karam, 1997).

At the classroom level, achievement will be influenced by such factors as: the use of year-round or traditional calendars, the willingness of schools to utilize combination grade classes to manage enrollments, and the extent to which students are segregated by socio-economic status, ethnicity, language fluency levels, ability, gender or special education category (e.g., Burns & Mason 1998; R. Mitchell & D. Mitchell 1999; Rowan & Miracle, 1983; Shields & Oberg, 2000; Veenman 1995). Any of these factors might obscure the effects of CSR.

Teacher assignments also vary from class to class. Confounded with class size reduction we are likely to find variations in teacher credentials, experience, age, contract status, ethnicity, gender and educational attainment (e.g., Alexander, Entwisle, & Thompson 1987; Darling-Hammond, 1998; Ogawa, Huston, & Stine, 1999; Wright, Horn, & Sanders 1997). Finally, school and district boundaries serve to segregate students by neighborhood, culture, socio-economic background and other factors that are not easily measured (e.g., Arum, 2000; Black,

1999; Clotfelter, 1998; Entwisle, Alexander, & Olson 1997). All of these factors need to be considered as possible sources of achievement variation before we can confidently conclude that students have benefited significantly from taking instruction in reduced size classes.

Third, while most attention is focused on the *average* level of achievement for all students experiencing smaller classes, it is not clear that this is the only or even the most important outcome of interest. CSR might be judged successful if it provided the benefits only to the children in greatest need of academic help; or it might be seen as a failure if it interfered with the achievement of specific groups.

If, for example, classroom averages remain relatively constant, but previously failing students are now meeting grade-level standards, would that suffice to justify the expense of this policy? Or, if class averages go up, but low attaining students are no better off than they were before, would that be considered a failure? If class averages go up, but the attainment of students is concentrated on the middle range, so that previously high attaining students are no longer moving ahead as rapidly, would that be considered a failure? In short, what *patterns* of classroom attainment are being generated, and how are those patterns to be evaluated?

Fourth, particularly in California, implementation procedures may have distorted the normal, long-term impact of CSR because schools had to find classroom space and new teachers on short notice in circumstances when both were in short supply (Bohrnstedt & Stecher, 1999; Hymon, 1997; Illig, 1997; Ogawa, Huston, & Stine, 1999; Stecher & Bohrnstedt, 2000; Wexler, et al., 1998). By the same token, if we put off assessing its impact until all implementation wrinkles are straightened out, it will be impossible to separate CSR from other factors affecting overall student achievement.

Since local school districts had to implement the policy in a matter of a few months, it was difficult to make needed changes in classroom space and teacher recruitment. Schools of education had no advanced warning, with the result that they prepared no surplus of new teachers to take up the large number of new teaching positions created. Construction companies did not have an opportunity to gear up for the production of new classroom facilities. Even if they did anticipate construction needs, there was no early release of construction funds to prepare classrooms. New teachers, not fully qualified teachers, and teachers transferring to new assignments at the last moment had to start instruction of smaller classes in new spaces. Sometimes such irregular spaces as libraries, multipurpose rooms or computer laboratories were converted for the new classes. A significant number of these problems continue into the second and subsequent years of implementation.

Fifth, since California does not require systematic achievement testing of students until the end of second grade, it is not possible to ascertain whether CSR in this state is having substantial impact during this first critical year of schooling. Results from Tennessee's Project STAR indicate that the major effects of class size reduction are experienced during the kindergarten year, or during the first year a child is exposed to this form of instruction (e.g., Finn & Achilles 1990; Finn, Gerber, Achilles, & Boyd-Zaharias, in press; Folger & Breda, 1989; Krueger 1999; Nye, Hedges, & Konstantopoulos, 2000). If this is generally true, it may not be possible to measure the effects of class size reduction in settings like California where the small class experiences could begin in the first, second or third grade and may not be encountered by some children until their second or third year of schooling. Additionally, it is possible, that achievement gains produced during an initial exposure to small classes will not be sustained over

time. Careful attention to this issue is required before the job of evaluation can be considered complete.

Study Design

This paper assesses the educational experiences of third and fourth grade students in seven Southern California school districts. The district enrollments range in size from about 600 to nearly 36,000 and represent a broad cross-section of urban, suburban and rural settings. The study design has four important features. First, the study is longitudinal in nature, examining the ultimate achievement levels of students whose individual class size histories are known. Second, the data analysis examines the CSR experiences of seven groups of third and fourth grade students who have had zero, one, two or three years of experience in small classes starting in first, second, or third grade. Third, perhaps the most important feature of the study design is that it allows us to estimate the confounding effects of a broad range of variables that could be masking the true effects of CSR experience. And fourth, analysis procedures recognize that key variables operate at four distinct levels: 1) individual and family factors, 2) classroom assignment variables, 3) teacher experience and demographic factors, and 4) school level variables.

The analyses presented here are based on carefully tracing the experiences of students in school districts where, due to implementation decisions made by district leaders, both large and small classes were created for children in all of the target grades (kindergarten through grade three), except kindergarten, in the first two years of implementation. All available records from students in regular classrooms (i.e., not community schools, individual tutorial students, special education Special Day Class classrooms, or combination grade classrooms with more than two grades) in each of the two study grades within each of the participating districts are included in

this study. They consist of 15,267 third and fourth graders in 697 classrooms in 74 schools. The student records selected for analysis are those for which a three-year history of class size reduction experience could be determined, where complete matching of students with teachers could be made, and where complete data on student classroom assignments were available. Of the original sample, 2,964 students were lost due to incomplete data, leaving 12,303. The largest portion of the sample reduction is due to incomplete class size reduction experience histories, which is almost entirely due to inter-district mobility resulting in discontinuities in student records. (We were able to secure records for most students who moved between schools within the same district, which is the only type of transient student retained in the analysis.)

The final sample for analysis consisted of 11,716 students with complete data and a total reading subject score, 12,039 with complete data and a total mathematics subject score, and 11,943 with complete data and a total language subject score. For the detailed multi-level analysis of total mathematics achievement presented here, the sample is further reduced to 11,262 students with completed data and a total mathematics subject score for both the "current" year (1999) and the previous year (1998). The reading and language total subject outcomes are not similarly analyzed. In addition to the lack of motivation due to insubstantial effects in reading and language achievement (see Findings), there is more sample loss due to missing scores in both of these subject areas from the 1998 testing cycle, making a value added analysis for reading and language far less representative.

The second major consideration of our study design is the identification of groups of students with contrasting CSR experiences. As indicated in Table 2, the study sample contains students with seven different patterns of exposure to reduced size classes. Among students currently in third grade, most (4,925) have had three years of CSR starting in first grade, a

Table 2. Class Size Reduction Experiences for Seven District Sample of 3rd and 4th Grade Students through the 1998-1999 School Year

		Current Grade in School		Total
		3	4	
CSR Experience	2 Years Starting 1st Grade	937		937
	3 Years Starting 1st Grade	4925		4925
	1 Year Starting 2nd Grade		1247	1247
	2 Year Starting 2nd Grade	469	801	1270
	1 Year Starting 3rd Grade		1783	1783
	None		2141	2141

substantial group (937) had two years of exposure starting in grade 1 (but returning to large classes in either second or third grade), and a moderate size group (469) had two years of exposure starting in grade 2.

Among fourth graders in the study sample, the largest group (2,141) had no CSR experience; none of the fourth graders had started their CSR exposure in grade 1. Substantial groups of fourth graders have had either one or two years of exposure starting in either second or third grade.

As the data analyses presented below will confirm, the timing of exposure to CSR is quite important. Students whose earliest CSR experience was in the first grade showed quite different results from those whose initial exposure was in second or third grades (no students with their initial exposure in kindergarten were available for this study). Exactly how important these differences are is hard to estimate because, by the third year of implementation, there were no students in the third grade who had spent no time in reduced size classes, and there were no students in the fourth grade who had started their CSR exposure in first grade.

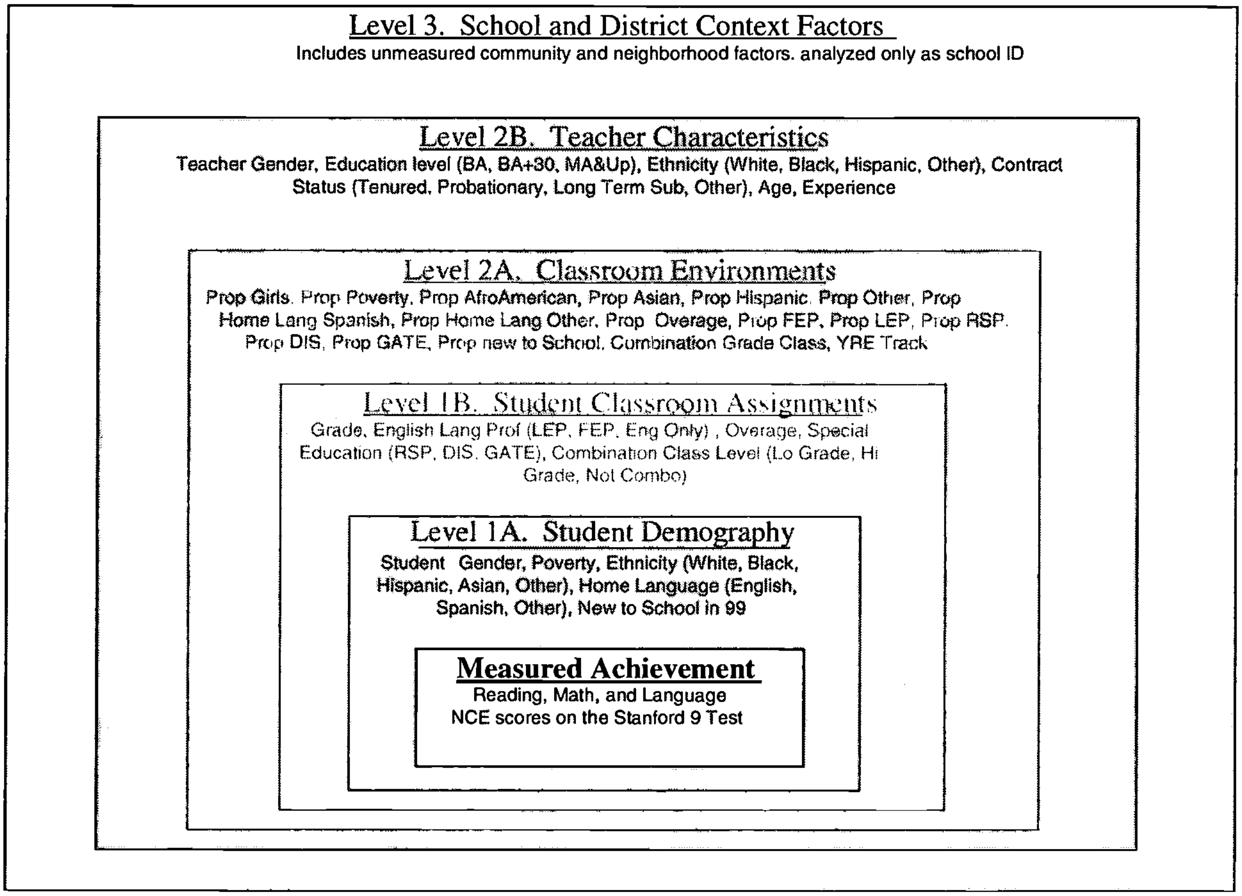
Importantly, our third study design feature, we are able to estimate the confounding effects of a broad range of variables that could be masking the true effects of CSR experience.

Since California's CSR initiative was implemented as a rapidly expanding, full-fledged operational program, practical considerations made it inevitable that the children placed in small classes would not have the same demographic profiles, classroom contexts or prior achievement histories as those who continued to attend school in larger classes. By monitoring their complete demographic profiles, their teacher characteristics and their school program assignments, the study is able to statistically control for these biasing factors and thus produce a much more accurate estimate of the true CSR impact on achievement.

Our fourth design feature is the explicit multi-level character of the data set. At the student level, we have individual student demographic factors, including family measures such as home language and income status, and school program participation identifiers such as special education and English language learner status. At the classroom level, in which students are "nested," we have both classroom characteristics and teacher characteristics. The kind of classroom to which a student is assigned is identified by its demographic and special program composition, attendance calendar, and combination grade status. The classroom teacher is distinguished by demographic and professional characteristics such as age, ethnicity, gender, teaching experience, contract, and credential. At the school level, in which classrooms are "nested," the composition of the student population, the teaching staff, and other features such as grade range and attendance calendar are available to include in the analysis. A complete model of the variables studied is presented in Figure I.

The specific variables are described in detail in Appendix A.

Figure I.



Validating the Study Sample

The students in the CERC study sample are quite representative of California's total public school student population. Table 3a presents a statistical comparison of the 12,303 students and their classroom teachers in our sample with the 947,597 California students in grades 3 and 4. As shown at the top of the table, the two groups are very closely aligned on overall achievement in reading, mathematics, and language. While the sample is generally representative of California's total school population, there are half a dozen places where the sample deviates substantially from the overall statewide population. As would be expected from the design of the study, which takes advantage of incomplete CSR implementation, the

proportion of students in reduced size classes is somewhat below the state as a whole. Our study sample has more English home language students than the overall state population, with commensurately fewer Spanish and Other home language students. Correspondingly, there are fewer English Learner (LEP) students. Despite the high number of students in the study sample from low-income homes (NSLP eligible), the California proportion statewide is yet higher. White and Hispanic student populations are fairly similar, but the sample has a higher proportion of African/Black students and a lower proportion of Asian students. The sample also has less than half of the state's proportion of its students attending traditional calendar schools. This reflects an increasing use of the multi-track year-round calendar to create classroom space for CSR among the sample districts.

Table 3a also presents some descriptive statistics for the study sample on variables for which statewide population parameters were not available at the time this report was prepared. About 13 percent of the sample students are in combination grade classes. Nearly one out of every eleven students was new to the school where they were tested in 1999. Among year-roundeducation tracks, Track C and Track D are the preferred ones. Together they enroll 21 percent more students than Tracks A and B. In our sample, there are only two year-round schools on 3-track attendance calendars, but the dates align perfectly in one case and nearly perfectly in another with three of the four tracks on the 4-track attendance calendars. As such, it is the attendance calendar for the track that determines each student's and teacher's designation.

Table 3b compares teacher characteristics in the CERC study sample with statewide averages. Teacher ethnic distribution reflects the student distribution reported in Table 3a. There are noticeably more African/Black teachers in the sample and fewer Other ethnicity

Table 3a. Comparison of Mean Achievement and Percentage of Students by Level for Each Factor for Seven District Sample with Elementary Schools Enrolling Third and Fourth Grade Students in the State of California in 1998-1999.

Factor	Levels	Sample	State
Mean SAT-9 Subject Total Achievement (NCE)	Reading	44.2	45.2
	Mathematics	48.3	48.3
	Language	47.2	47.2
Grade	3	51.3	51.5
	4	48.7	48.5
CSR Option 1 in 1996-97 (Grades 1-2)	Yes	63.9	71.5
	No	36.1	28.5
CSR Option 1 in 1997-98 (Grades 2-3)	Yes	72.5	80.7
	No	27.5	19.3
CSR Option 1 in 1998-99 (Grades 3-4)	Yes	44.2	44.0
	No	55.8	56.0
Student Ethnicity	African/Black	14.2	9.0
	Asian	3.2	7.5
	Hispanic	45.2	42.7
	White	35.1	36.8
	Other	2.3	3.9
Student Home Language	Spanish	23.5	31.0
	English	73.3	60.1
	Other	3.2	8.9
Student Low Income Status (NSLP Qualified)	Yes	47.1	56.0
	No	52.9	44.1
Student Gender	Male	50.6	51.0
	Female	49.4	49.0
Student Intra-District Mobility 1998-1999	New to School	9.3	N/A
	Not New	90.7	
Student English Language Proficiency	LEP	16.7	30.2
	FEP	10.0	9.7
	English Only	73.3	60.1
Student Special Education/ GATE	RSP	3.8	N/A
	DIS	2.1	
	GATE	9.8	
	Not Spec Educ	84.2	
Student Overage for Grade (15+ Months)	Overage	2.4	N/A
	Not Overage	97.6	
Student Grade in Combination Grade Classroom	Low Grade	7.0	N/A
	High Grade	5.7	
	Single Grade	87.3	
School Attendance Calendar	Traditional	41.9	86.6
	YRE A-Track	12.9	12.4
	YRE B-Track	13.3	
	YRE C-Track	16.1	
	YRE D-Track	15.7	
	YRE Single Track	0.0	1.1

teachers. The sample has an appreciably higher percentage of male elementary teachers for students in grades three and four than the overall state percentage for schools enrolling studentsin grades three and four. The proportion of fully credentialed teachers is only slightly higher than that of the state as a whole. The sample also has less than 15 percent more probationary teachers than the state population, matched by a reduction in the number of teachers having tenure contracts. Though the distribution differs, the total number of teacher on "temporary" or "other" contracts is nearly the same. There are a few percent more teachers with 30 semester hours beyond the bachelor's degree, matched by a reduction in those holding just the bachelor's

Table 3b. Comparison of Percentage of Teachers by Level for Each Factor and Average Teaching Experience for Seven District Sample with Elementary Schools Enrolling Third and Fourth Grade Students in the State of California in 1998-1999.

Factor	Levels	Sample	State
Teacher Ethnicity	African/Black	7.6	4.9
	Hispanic	12.1	14.3
	White	77.3	74.1
	Other	3.0	6.7
Teacher Gender	Male	20.8	14.7
	Female	79.2	85.3
Teacher Credential Status	Full Credential	88.1	86.6
	Not Full Credential	11.9	13.4
Teacher Contract Status	Long-Term Sub/Temp	2.0	8.0
	Probationary	23.0	20.4
	Tenure	62.3	64.7
	Other	12.7	7.0
Teacher Education Level	MA & Up	25.6	26.1
	BA + 30	55.2	51.3
	BA	19.2	22.7
Avg. Teacher Experience (Years)		10.2	11.9

degree. Finally, the average teacher experience level for the grades three and four classrooms in this sample is more than a year-and-a-half lower than that for the state.

The final dataset was produced by combining SAT-9 data with CBEDS-PAIF data and retaining for study all students in grades three and four on whom it was possible to document their entire history of CSR participation. These data, plus information on district CSR implementation generated the fifteen control variables described in Appendix A. Calculating classroom and school averages for these variables created an additional 16 context and control variables.

As described in the design section above, the final sample consists of seven groups of students with differing combinations of starting grade and duration of exposure to CSR (see Table 2). Two tiny groups of very exceptional students were dropped from the study because they consisted of retained students or those taking fourth grade instruction in a small 3-4-combination grade class.

Data Analysis

Data analysis for this study utilized three multivariate techniques.[1] Multiple Discriminant Analysis (MDA) was used to examine the issue of implementation bias, documenting the extent to which California's non-experimental program strategy confounded CSR exposure with other important factors that could be expected to influence achievement. This was done at the classroom level, the level at which the policy is implemented. In particular, membership in the seven CSR experience groups was predicted by measures of classroom composition, attendance calendar, multi-grade status, and classroom teacher characteristics. Discriminant analysis

[1] *SPSS for Windows 9.0* was used for discriminant, ANOVA, and OLS regression analyses (SPSS, Inc., 1999). *HLM for Windows 5.02* was used for multi-level regression analyses (Raudenbush, Bryk, & Congdon, 2000).

provides a test of the implementation bias hypothesis (Hypothesis 1) by identifying the extent to which student CSR experiences are associated with particular classroom characteristics.

Once implementation bias was documented, ordinary least squares (OLS) regression and two- and three-level Hierarchical Linear Modeling (HLM) were used, as appropriate, to separate student-, classroom- and school-level influences on achievement. Initially, all achievement effects were examined using HLM. When higher level variance components in the unconditional model were found to be statistically not significant so that only the lowest unit of analysis had significant variation, which only occurred for the analysis of the dependence of classroom kurtosis, then OLS regression was employed for hypothesis testing.

There are three major benefits of HLM over OLS regression that motivate contending with its complexity for this analysis (discussions of these and other points are found in Bryk & Raudenbush, 1992; Goldstein, 1995; Snijders & Bosker, 1999). First, when there is significant variance at more than one level (e.g., within classroom [the student level] and between classrooms within schools [the classroom level] or between schools [the school level]) the standard errors for the regression coefficients at the higher levels (classrooms and schools) are more accurately estimated using HLM. This is critical to the acceptance or rejection of a hypothesis test. With OLS regression, the standard errors are often terribly underestimated for classroom and school level effects, leading to the conclusion that statistically significant effects have been estimated. If only one level in the multi-level or nested model has a significant variance components (e.g., there is only significant variation at the student level while the variance components at the classroom and school levels are not significantly different from zero) then OLS regression at the one level is a simpler, if not superior methodological choice.

The second major benefit to HLM is that observation within units, i.e., classrooms and schools, are allowed to be correlated without violating the assumption of the statistical method being employed. This is possible by mixed effects modeling, i.e., both fixed and random effects. The random effects are allowed to covary with each other as well as provide the hierarchical variance components. Since students have the same teacher within a single classroom at the elementary level in most schools (the self-contained, multi-subject classroom), students outcomes are expected to be correlated because of this unitary influence. Achievement is modeled as having a between classrooms (within schools) random effect. Similarly, many educational opportunities are structured by school policies, personnel, and resources, especially the neighborhood attendance area, leading to the expectation that classroom outcomes within schools would be correlated. Achievement is further modeled as having a between schools random effect. Utilizing HLM instead of OLS regression allows the analyst to proceed without being concerned about this violation of independent observations within units of aggregation (classrooms and schools) by treating the random effect at each aggregate unit as being drawn from random distribution (of classrooms and schools) with the sample grand mean as the estimate of the central tendency of these distributed effects.

The third benefit to HLM important here is purely technical. The methodological advance being exploited here is that we are not required to have a balanced nested design. Before the modern computational algorithms were put to use, equal number of observations were required for each cluster (i.e., the same number of students for each classroom and the same number of classrooms for each school). This old requirement stands in direct opposition to the phenomenon under investigation, the effect of different class sizes, and does not easily

accommodate the dramatic variation in the number of classes within schools that comes from the fact that we have both urban and rural districts in the sample.

Hypotheses 2 through 6 explicitly require multi-level hypothesis testing using HLM (see Appendix B for model details). After reviewing results from an earlier block entry OLS regression analysis (D. Mitchell & R. Mitchell, 2001), the overall impact hypothesis (Hypothesis 2) is tested using a three-level model, where class size reduction effects are specified at the student level for prior CSR experience with current CSR experience at the classroom level. Prior CSR experience is modeled using four dummy-coded variables at the student level indicating whether students had (coded as 1) one year of previous CSR experience, two years of previous CSR experience, started their CSR experiences in first grade, started their CSR experiences in second grade, or not (coded as 0). Current CSR experience is modeled using one dummy-coded variable at the classroom level indicating whether students were in a reduced size classroom (coded as 1) or not (coded as 0) in the "current" testing year. Current CSR experience is modeled as having a school level random component (i.e., the within school effect of CSR is allowed to vary from school to school). Except for prior achievement, no student-level coefficients depend on current CSR experience. Prior achievement is modeled as having a between classroom and between school random component. That is, the impact of prior achievement varies from classroom to classroom within schools as well as from school to school. In addition to a long list of implementation bias control variables, current and prior achievement pattern variables are entered into the model to test Hypothesis 2. Prior achievement at the individual student level is not entered because any capitalized effects of early class size experiences would be removed by such a specification. This model provides a test of whether or not CSR has a reliable overall impact on student achievement.

The socialization hypothesis (Hypothesis 3) is tested using a three-level model, where in addition to indicating that students started their CSR experiences in first grade at the student level, the interactions of starting CSR in first grade with student race/ethnicity (dummy-coded levels: African American, Asian American, Hispanic, and Other Non-White), home language (dummy coded levels: Spanish and Other Non-English), and family income status (dummy-coded levels: Free Lunch Qualified and Reduced Price Lunch Qualified) are entered. Prior achievement and classroom pattern variables are not included in the model because both individual and collective benefits of early class size experience may be capitalized, leaving no value-added effect to be detected. As in all cases, implementation bias control factors are specified. This model provides a test of whether there were achievement benefits from starting in first grade, and whether there were additional benefits for those students more likely to require additional early school socialization.

The classroom management hypothesis (Hypothesis 4) is tested using a three-level model, where in addition to indicating that students are currently in a reduced size classroom or not, the interactions of being in a reduced size class with minimum number of students (3) in the classroom of a particular race/ethnicity (dummy-coded threshold levels: African American, Asian American, Hispanic, and Other Non-White), home language (dummy coded threshold levels: Spanish and Other Non-English), family income status (dummy-coded threshold level: Free or Reduced Price Lunch Qualified), or special education status (dummy-coded threshold levels: DIS and RSP) are entered. Additionally, prior achievement is included with the implementation bias control factors to isolate current effects in the currently reduced size class – to determine the value added from the current experience. This model provides a test of whether there were achievement benefits from being in a reduced size class for students who were in

classes that more likely would be difficult to manage than if there were fewer than the minimum threshold of academically at-risk students.

The school resources hypothesis (Hypothesis 5) is tested using a three-level model, where in addition to indicating whether students are currently in a reduced size classroom or not, the interactions of being in a reduced size class with school level circumstances associated with resource challenged schools (i.e., the proxy variables are school proportion of students by levels of race/ethnicity and poverty, school proportion of fully certified teachers, and school average teacher experience level) are entered. As with the other models, implementation bias control factors are specified. Two separate models are considered. First, prior achievement is excluded because it is possible that the marginal additional return to class size reduction may be small compared to the total benefit over all years of CSR exposure. However, since resource utilization should not necessarily depend on timing of treatment, the value-added marginal return to a current class size reduction as a function of school resource proxies is specified by controlling of prior achievement. This model provides a test of whether students in a reduced size class accrue an additional achievement benefit from improved resource utilization in resource disadvantaged schools.

The instructional practices hypothesis (Hypothesis 6) is tested using a set of two-level models, where in addition to indicating whether the classroom is a reduced size class, prior achievement classroom pattern variable are entered – the first four moments for the distribution of prior achievement in the class (i.e., the prior achievement mean, standard deviation, skewness, and kurtosis). For this set of hypotheses tests at the classroom level, the dependent variables are the current classroom pattern variables, each tested separately. That is, four models, each with the same set of independent variables, are tested: the current class mean as outcome, the current

class standard deviation as outcome, the current class skewness as outcome, and the current class kurtosis as outcome. Similar to the previous models, a set of classroom level implementation bias control factors is also specified. These models provide a test of whether teachers produce an instructional environment that alters the patterns of student achievement in their classrooms. For example, they may raise the mean – shift the central tendency of student performance – lower the standard deviation – bring students closer to the same outcome level – raise the skewness – bring the low performing students toward the modal outcome – or lower kurtosis – suppress the likelihood that there will be many if any extremely high or low performing students in the class.

Findings

Findings resulting from the data analysis described above can be easily summarized in terms of the six basic hypotheses outlined earlier in this paper.

Finding #1: <u>The Implementation Bias Hypothesis</u> – Bias in CSR implementation is large, potentially confounding all CSR effects

CSR implementation provided different groups of students with very different types of exposure to smaller classes, making statistical control over a wide variety of confounding variables absolutely essential if we are to discover the true effects of small class exposure on achievement. Table 4 summarizes some of the most obviously confounded variables that could easily obscure the impact of CSR on student achievement. Each of the factors in this list is significantly related to achievement and they interact with each other in complex and sometimes unpredictable ways.

Each of the variables in Table 4 was measured at the classroom level. For all of the teacher variables except experience, the measure is a dummy coded scoring of whether the teacher did (1) or did not (0) have the characteristic. Experience was measured in total years as a teacher. All the student level variables were measured as the percentage of the class having each

Table 4. Substantial Contributors to Multiple Discriminant Analysis of Implmentation Biases in Seven CSR Exposure Groups

Variables that Substantially and Significantly Discriminate Among CSR Exposure Groups (All Univariate ANOVA's are significant at p < .001)											
	Class Size Experience (Current Grade, Time, Start Grade)										
Current Grade:	In 4th	In 3rd	In 3rd	In 4th	In 3rd	In 4th	In 4th	Average	Range		Pct Var
Teacher Variables CSR Exper	None	2Yrs-1st	3Yrs-1st	1Yr-2nd	2Yrs-2nd	2Yrs-2nd	1Yr-3rd	All Groups	Min	Max	Explained
Teachers: Other Credential	25%	5%	14%	14%	0%	0%	1%	12%	0%	25%	6%
Teachers: Probationary Contract	15%	44%	19%	22%	14%	41%	26%	23%	14%	44%	4%
Teachers: Long Term Sub Contract	3%	5%	1%	2%	0%	0%	1%	2%	0%	5%	1%
Teachers: Years Experience	9.6	6.6	10.8	10.1	15.7	11.7	10.6	10.4	6.6	15.7	3%
Teachers: Not fully credentialed	12%	20%	11%	12%	1%	7%	8%	11%	1%	20%	1%
Class Ethnic Composition											
Class: Pct Afro American	15%	25%	15%	21%	8%	23%	10%	16%	8%	25%	9%
Class: Pct Other Ethnicity	2%	3%	2%	2%	5%	3%	4%	2%	2%	5%	5%
Class: Pct Asian	2%	2%	3%	3%	6%	3%	4%	3%	2%	6%	4%
Class: Pct Hispanic	44%	52%	46%	45%	31%	46%	40%	45%	31%	52%	3%
Class Language Status											
Class: Pct LEP	19%	18%	17%	14%	11%	11%	15%	16%	11%	19%	1%
Class: Pct FEP	7%	10%	10%	11%	6%	14%	7%	9%	6%	14%	5%
Class Special Education											
Class: Pct RSP Pgm	5%	5%	3%	6%	4%	6%	6%	5%	3%	6%	4%
Class: Pct DIS Pgm	2%	2%	2%	2%	0%	1%	1%	2%	0%	2%	4%
Class: Pct GATE Pgm	10%	8%	7%	9%	16%	8%	12%	9%	7%	16%	2%
Class Composition											
Class: Pct New	17%	27%	19%	24%	16%	34%	19%	21%	16%	34%	6%
Class: Pct Overage	2%	4%	2%	2%	4%	3%	4%	3%	2%	4%	4%
Class: Pct Girls	49%	49%	49%	50%	55%	50%	50%	49%	49%	55%	2%
Class: Combo Class	15%	19%	9%	18%	24%	8%	13%	13%	8%	24%	2%
Class: Pct Poverty	48%	47%	45%	48%	38%	40%	42%	45%	38%	48%	1%

of the identified characteristics. Thus, these variables measure the classroom context within which CSR implementation has taken place rather than the characteristics of the individual students being exposed. Though all show statistically reliable differences across the seven different class size experience groups, some are much more deeply entangled in CSR implementation than others. The nineteen variables in this table are clustered into groups for easier interpretation. The most powerful variable (accounting for about 9 percent of the variance in CSR exposure group membership) is the percent of the students in the class that are African American. The least powerful, but still highly reliable predictor variable is the percent of the class designated as Limited English Proficient.

Some of the variables strongly associated with CSR implementation are not very powerfully linked to achievement (like teacher experience). Other variables, not strongly linked to CSR implementation (like the class percentage of Poverty students) exert a lot of influence on achievement. Poverty only predicts about 1 percent of the variance in CSR implementation, but

it has a powerful effect on achievement. Home language (not shown on the table) and participation in GATE programs are also strongly associated with achievement but only weakly associated with CSR implementation.

Unfortunately, given the modest size of CSR impact (which we will discuss at more length in a moment), even moderate confounding with CSR can mean substantial differences between groups of children with different CSR exposure. Poverty rates, for example, range from a low of 38% of the third-grade students in classrooms where children had two years of CSR starting in the second grade to a high of 48% of those who had no CSR exposure or who had only one year starting in the second grade. A ten percent difference in classroom poverty rates could well produce a negative effect on achievement that could fully offset any gains being produced by a year of participation in small class instruction.

Average teacher experience varies by more than 9 years. The percent Probationary Contract teachers varied dramatically – from 14 percent for third-grade children getting two years of CSR starting in the second grade to 44 percent for the third-graders who got two years of CSR starting in the first grade. The Limited English Proficient student proportion varied from less than 11 percent to over 19 percent for students with no CSR experience.

Examples could be multiplied endlessly here. The basic point is that implementation of CSR deeply entangled with other variables known to influence student achievement. Taken together, these confounding variables have a multiple correlation squared of more than .54. That is, they predict about 54 percent of the variance in the type of CSR experience children have had – making it abundantly clear that they are at least as likely to be the causes of achievement variations among CSR implementation groups as are the small classes themselves.

The only sensible way to proceed is to remove the effects of these confounding variables ***before*** trying to assess the impact of CSR. This is done using a statistical regression procedure that, in effect, equalizes the different CSR treatment groups on these variables before testing to see whether the groups, so conditioned, have significantly different achievement test scores.

Finding #2: ***<u>The Overall Impact Hypothesis</u>*** ***– Overall, CSR impact on achievement is quite modest***

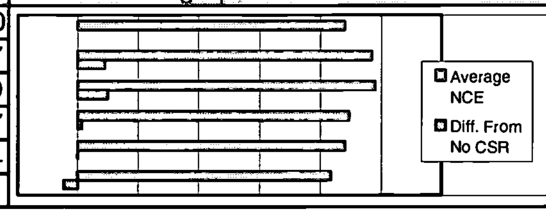
Finding substantial implementation bias, we set about to statistically control the biases in order to isolate the true impact of CSR experience on academic achievement. Before trying to map the complex relationships between CSR and the other nested variables in the study, we used Block-Entry Multiple Regression analysis to get a rough and ready assessment of whether CSR has an easily identifiable impact on achievement or is so small in effect at to require the most rigorous scrutiny of potential biasing factors. In our report on data collected at the end of the second year of California's CSR implementation (D. Mitchell & R. Mitchell, 1999), we found that the effect was very small and quite unstable in the first year of implementation, and almost as small in the data for year two. The next three tables present the final outcomes of the block-entry regression analysis on mathematics, reading and language achievement test scores for the third year of implementation.

Jumping right to the infamous "bottom line," Table 5 reports the relative mathematics achievement gains for students with various combinations of CSR when compared to the 2,141 students in the study sample who had no CSR experience at all. The estimated achievement levels reported in the table are those that would be expected if all demographic, classroom assignment, classroom environment, teacher characteristics and school and district effects are statistically equalized for all students in the study group.

Table 5. Average SAT-9 Mathematics Achievement by Class Size Reduction Experience
(NCE scores adjusted for all known implementation biases)

Starting Grade	Number of Years	Average Math Score	Difference from No CSR
No CSR	Zero	43.95	0
First	Two	48.42	4.47
First	Three	48.94	4.99
Second	One	44.62	0.67
Second	Two	43.99	0.04
Third	One	41.64	-2.31

Bargraph of Test Scores



The table reports the average SAT-9 mathematics achievement for each CSR exposure group. Achievement is reported in Normal Curve Equivalent (NCE) scores, which have a nationally normed mean of 50, standard deviation of 21.06. (A change of about 10 points represents one year of academic progress – this number varies from one grade to another). The actual mean for our sample was 48.3, a bit below the national mean but right in line with the California state mean. The standard deviation for our sample was 20.96, quite close to the expected value.

As shown at the top of the first column of numbers in the table, the estimated average for students who had no CSR exposure was 43.95, well below the 48.3 overall average. The other numbers in this column of the table report the average achievement for the groups of students with each of the five different CSR experiences. Only those students who had their initial exposure to small classes during their first grade in school show any significant improvement in their mathematics achievement. Those who started CSR in the second grade scored virtually the same as students with no CSR experience, and those whose first exposure was in grade three actually did less well than if they had received no CSR exposure at all. Indeed, the average of all the CSR exposed student groups would be only 1.57 NCE points above the students with no CSR experience – about 5 weeks of normal academic progress.

The four and a half to five point advantage attained by the first grade exposure groups represents nearly a half year of an academic year advantage over their no CSR peers. This advantage, if it can be believed, compares favorably with that documented in Tennessee's Project STAR. Unfortunately, we must urge extreme caution in accepting this finding as definitive. All of the students in the sample which having no CSR experience were in the fourth grade at the time of this study, and all of the students receiving their first CSR exposure in the first grade were in the third grade when the data were collected. Thus, the differences between these groups could be due to an age-cohort difference between the third and fourth grade students. Nevertheless, the differences are significant and in favor of early exposure to small classes. By the end of the fourth year of CSR implementation we will be able to determine whether the effects are reliably related to CSR experience.

Tables 6 and 7 present the same information for reading and language achievement as was presented for mathematics in Table 5. Here we see that, when the same equalization procedures are applied to achievement in reading and language, the effects of any type of exposure to CSR are very small in size and mixed in direction. The most positive benefits (though extremely small) positive effects are still concentrated on students who start CSR earlier

Table 6. Average SAT-9 Reading Achievement by Class Size Reduction Experience

(NCE scores adjusted for all known implementation biases)

Starting Grade	Number of Years	Average Read. Score	Difference from No CSR	Bargraph of Test Scores
No CSR	Zero	43.15	0	
First	Two	43.79	0.64	
First	Three	43.46	0.31	
Second	One	43.77	0.62	
Second	Two	42.36	-0.79	
Third	One	41.96	-1.19	

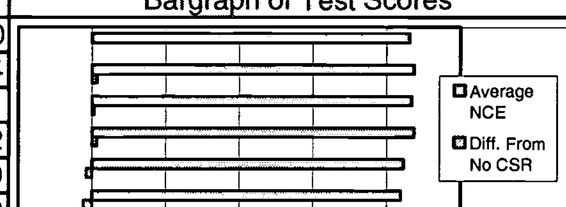
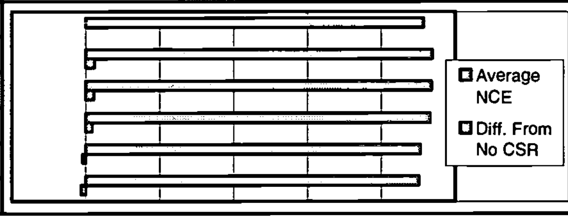


Table 7. Average SAT-9 Language Achievement by Class Size Reduction Experience
(NCE scores adjusted for all known implementation biases)

Starting Grade	Number of Years	Average Lang. Score	Difference from No CSR
No CSR	Zero	45.73	0
First	Two	46.9	1.17
First	Three	46.74	1.01
Second	One	46.55	0.82
Second	Two	45.27	-0.46
Third	One	45.05	-0.68

Bargraph of Test Scores



and persist in the small classes longer. While some of the differences on this table are statistically reliable, when compared with the effects of other variables (discussed in the next section of this report) they appear truly trivial in magnitude.

The language scores reported in Table 7 present a pattern nearly identical to that for reading. Very small positive benefits for children who started CSR in first grade, with no benefit or possibly slight losses in achievement for students who begin their CSR exposure later.

Table 8. Initial HLM Estimates of Class Size Reduction Achievement Effects for Seven District Sample of 3rd and 4th Grade Students through the 1998-1999 School Year (Third Year of CSR Implementation).

		Current Grade in School: 3	Current Grade in School: 4	Mean
CSR Experience	2 Years Starting 1st Grade	3.08		3.08
	3 Years Starting 1st Grade	*4.97*		4.97
	1 Year Starting 2nd Grade		0.67	0.67
	2 Year Starting 2nd Grade	*2.56*	0.24	1.05
	1 Year Starting 3rd Grade		-0.42	-0.42
	None		0.00	0.00

Note: *Italicized* cells values are for students current in reduced size classes.
N=11,262.

Since mathematics achievement scores point to a probable impact of California's approach to CSR on student achievement, the rest of our analysis will concentrate on

documenting the extent a possible reasons for that impact. In Table 8 we present a Hierarchical Linear Model (HLM) estimation of the same CSR impacts shown in Table 5 (i.e. the extent to which CSR appears to affect mathematics achievement when the effects of all of our known biasing factors are statistically controlled at the appropriate level).

The HLM analysis presents a similar, but slightly rosier picture of CSR impacts on mathematics achievement. The largest estimated improvement, 4.97, is for third graders who have had three years of CSR starting in grade one. And the apparent significant loss in achievement for fourth grade students who had only a year of CSR in their third grade year is seen in the more sophisticated HLM analysis to have been not quite accurate (HLM estimates a slightly negative –0.42 NCE points for this group). The HLM approach estimated separately the third and the fourth graders who had two years of CSR starting in the 2nd grade. Those in the third grade, and therefore still in a small class at the time of testing show a 2.56 point gain for their CSR experience, but the fourth graders who are now in a large class show only a very small gain for their experience (0.24 NCE points). This analysis continues to support the view that students whose CSR experience comes earlier in their school career benefit the most. First and second grade starters grade starters get three to twelve times the benefit of the third grade starters.

Table #. Hypothesis 2 (Duration): Three-Level HLM for the Relationship of Years of Class Size Reduction Experience to the Achievement of Academically At-Risk Student Groups.

Predictors of Total Mathematics Achievement (SAT-9 Spring 1999)		Unstandardized Coefficient	Standard Error	*p*
Years of CSR Experience Effect		0.91	0.54	0.093
Years of CSR Experience by Student Identified as:	African American	-0.25	0.54	0.635
	Asian American	-0.40	0.94	0.673
	Hispanic	0.03	0.39	0.949
	Other Non-White	-0.55	1.07	0.607
	Spanish Home Language	-0.17	0.44	0.688
	Other Non-English Home Language	0.86	0.95	0.368
	Low Income/Free Lunch Qualified	-0.06	0.30	0.838
	Low Income/Reduced Price Lunch	-0.12	0.43	0.783
		Δ Deviance	Δ *df*	*p*
	Model Change:	59.18	12	0.000

NOTE: Model change statistics (students within classrooms within schools) were calculated by comparing a base model that controls for implementation biases with a model that additionally enters years of CSR experience, controlling for current CSR experience, and the eight student-level years of CSR interaction terms; since the Deviance has a χ^2 distribution, a one-tail χ^2 test was used to obtain *p*.

Table 9 provides an overview of just how powerful the variables used in this study are in predicting student, classroom and school level achievement for third and fourth grade students. This table summarizes the output statistics for five HLM models aimed at providing increasingly complete explanations of mathematics achievement variance. The first row of this table presents the total variance at each of the three levels in the model (students within classrooms, classrooms within schools, and between schools). These numbers show that the student level variance is 71

Table 9. The Explanatory Power of Hierarchical Linear Modeling: Remaining Variance at Each Level in a Three-Level Hierarchical Linear Model (within Classrooms, between Classrooms within Schools, and between Schools) as Explanatory Variables are Added to Account for Student Mathematics Achievement (SAT-9 NCE Scores from Spring 1999)

	Variance Components			Full Maximum Likelihood	
	Classrooms	Schools			
Three-Level HLM Models	Within	Within	Between	Deviance	*df*
1. Unconditional: (Total Variance [100%] at Each Level)	316.95	78.24	48.82	97110.40	4
2. Remove: Pupil & Classroom Biases	76%	51%	22%	93750.41	43
3. Remove: Class Size Reduction Experiences	76%	41%	24%	93686.49	49
4. Remove: Prior Achievement (Spring 1998)	37%	38%	12% [a]	86077.15	56
5. Remove: Classroom Pattern Variables	37%	34%	9% [b]	86027.23	63

NOTE: Additional variance components were modeled: Within schools component for prior achievement and between schools components for prior achievement and current class size reduction experience (small class in 1999). For all variance components, tabulated and untabulated, except where noted for between schools, $p < 0.01$.

[a] $p = 0.102$

[b] $p = 0.187$

percent of the total variance in achievement, only 18 percent between classes and 11 percent between schools. As the variables available for study are entered into the HLM models, progressively more of the variance is explained so that by the time classroom achievement pattern variables are controlled in model five, fully 67 percent of the achievement total variance has been explained; more that 90 percent of the school to school variations.

Finding #3: ***<u>The Socialization Hypothesis</u> – The hypothesis that CSR in the earliest grades more effectively socializes children to school is supported, but differentially benefits socialization for poor and underrepresented minority students is not supported.***

Table 10 presents the results for the block of variables entered into a three-level HLM analysis to test the socialization hypothesis. The unstandardized regression coefficient, its standard error and *p* value (based on the t-ratio) are reported in the three columns on the right side of the table for the main effect of beginning CSR in the first grade and for its interaction with eight student level demographic characteristics. The change in the Deviance statistic,

reported at the bottom of the table, is highly statistically significant for this test of Hypothesis 3 indicating that at least one of the variables related to beginning CSR in the first grade makes a

Table 10. Hypothesis 3 (Socialization): Three-Level HLM for the Relationship of Early Grade Class Size Reduction Experience to the Achievement of Academically At-Risk Student Groups.

Predictors of Total Mathematics Achievement (SAT-9 Spring 1999)		Unstandardized Coefficient	Standard Error	*p*
Begin CSR in First Grade Effect		4.84	0.96	0.000
Begin CSR in First Grade by Student Identified as:	African American	0.05	1.31	0.968
	Asian American	-0.69	2.33	0.766
	Hispanic	-0.21	1.08	0.845
	Other Non-White	-0.58	2.69	0.830
	Spanish Home Language	-0.50	1.23	0.683
	Other Non-English Home Language	1.67	2.43	0.492
	Low Income/Free Lunch Qualified	-0.02	0.75	0.984
	Low Income/Reduced Price Lunch	0.73	1.12	0.512
Begin CSR in Second Grade Effect		0.73	1.12	0.513
Begin CSR in Second Grade by Student Identified as:	African American	1.12	1.35	0.408
	Asian American	-1.00	2.86	0.725
	Hispanic	0.41	1.01	0.688
	Other Non-White	0.77	2.56	0.763
	Spanish Home Language	-0.07	1.28	0.954
	Other Non-English Home Language	2.51	2.87	0.381
	Low Income/Free Lunch Qualified	0.02	0.98	0.980
	Low Income/Reduced Price Lunch	-1.13	1.36	0.407
		Δ Deviance	Δ *df*	*p*
	Model Change:	60.01	18	0.000

NOTE: Model change statistics (students within classrooms within schools) were calculated by comparing a base model that controls for implementation biases with a model that additionally enters first grade CSR experience and the eight student-level first grade CSR interaction terms; since the Deviance has a χ^2 distribution, a one-tail χ^2 test was used to obtain *p*.

significant improvement in the model explaining student mathematics achievement after controlling for implementation biases and individual student characteristics.

Students who began their CSR experiences in first grade are attaining significantly higher mathematics achievement in the third grade than those students in third or fourth grade whose first CSR experience started later or had no CSR experience at all. At a level of 4.47 NCE

points, the benefit of CSR in the first grade approaches half of a year's advantage in achievement for the students who are currently in the third grade. This is noteworthy as well as statistically significant. This main effect for first grade exposure supports the proposition that reduced size classes in the earliest grades socialize students so that they do better in school.

None of the interactions of individual student characteristics with beginning CSR in the first grade are significant, however. All of the coefficients have p values greater than 0.25. There are no differential socialization benefits for students who are most often at-risk academically and who are likely to come from homes that are not culturally aligned with school expectations (i.e., "under-represented" racial/ethnic minority students, students from non-English speaking homes, and students from families near or below the poverty line). In this sample, early small class experiences are not providing additional benefits to at-risk students, at least not achievement benefits that are still detectable when these students reach the third grade.

Finding #4: <u>The Classroom Management Hypothesis</u> – The hypothesis that CSR eases classroom management is supported, but differentially benefits for more challenging classrooms is not supported.

Table 11 summarizes the results of the three-level HLM testing the classroom management hypothesis. It is structured identically to Table 10. The statistics reported are for the main effect of being in a currently reduced size class (Current CSR Experience) and for its interaction with nine classroom level demographic characteristics. In addition to race/ethnicity, non-English home language, and poverty as conditions that may contribute to making a classroom more difficult to manage, the instructionally challenging DIS and RSP special education categories are included in this hypothesis test. Since this is a classroom-level test, the interaction term coefficients are for the effects having at least three "challenging" students in a reduced size classroom and not for the effect on individual students. The change in the Deviance

statistic, reported at the bottom of the table, is highly statistically significant for this test of Hypothesis 4 indicating that at least one of the variables related to current CSR experience makes a significant improvement in the model explaining student mathematics achievement after controlling for implementation biases and individual student characteristics and prior achievement. That is, the value added to individual student achievement by current exposure to CSR is significant

Table 11. Hypothesis 4 (Classroom Management): Three-Level HLM for the Relationship of Class Size Reduction to Student Achievement in a Classroom with Management Challenging Classroom Composition.

Predictors of Total Mathematics Achievement (SAT-9 Spring 1999)		Unstandardized Coefficient	Standard Error	*p*
Current CSR Experience Effect		6.84	2.43	0.005
Current CSR by Class Proportion of Students Identified as:	DIS	20.04	16.86	0.235
	RSP	0.01	9.88	0.999
	African American	-6.53	5.14	0.204
	Asian American	-6.82	13.16	0.604
	Hispanic	-8.88	5.88	0.131
	Other Non-White	-9.36	13.31	0.482
	Spanish Home Language	5.05	3.93	0.200
	Other Non-English Home Language	13.15	12.23	0.282
	Low Income/Poverty	0.53	3.11	0.865
		Δ Deviance	Δ *df*	*p*
	Model Change:	71.20	13	0.000

NOTE: Model change statistics (students within classrooms within schools) were calculated by comparing a base model that controls for implementation biases and prior achievement with a model that additionally enters current CSR experience and the nine classroom-level CSR interaction terms; since the Deviance has a χ^2 distribution, a one-tail χ^2 test was used to obtain *p*.

Students who are currently experiencing CSR, these are only third grade students, are attaining significantly higher value-added mathematics achievement than those students not currently in a reduced size class. This supports the proposition that a reduced size class eases the burden of classroom management, resulting in higher student achievement. At a level of 3.63 NCE points, the benefit of current CSR experience approaches a third of a year's advantage in

achievement for the students who not are currently in a reduced size class. This is large enough to take seriously, but small enough to be at risk of decaying rapidly when treatment ends. That is, small achievement impacts are generally not robust. They have little staying power.

Similar to the socialization hypothesis, none of the interactions of classroom student characteristics with current CSR experience are significant. With the exception of classes that exceed the minimum threshold for poverty student, all of the coefficients have p values greater than 0.33, and this exception still exceeds the generally accepted minimum standard of $p<0.05$. Thus, there are no differential classroom management benefits in classrooms where there are more than just of couple of students who are at-risk academically or who come from homes that are not culturally aligned with school expectations (i.e., "under-represented" racial/ethnic minority students, students from non-English speaking homes, and students from families near or below the poverty line, and special education students). In this sample, the small class experience does not provide additional benefits to classrooms with at-risk students, at least not achievement benefits that are still detectable when these students are in the third grade.

Finding #5: The Resource Effectiveness Hypothesis – The hypothesis that CSR supports more effective use of school and classroom resources is not supported.

Table 12 presents a summary of the HLM analysis testing hypothesis 5 and is structured the same as Tables 10 and 11. This time, the added predictors are school level predictors of the effect of current CSR effect and not predictors of mathematics achievement itself. That is, within the three-level model of student achievement, there is a model specifying that the current effect of classroom level CSR is predicted by school level measures that serve as proxy measures for resource challenged schools. As this table shows, we could find no special benefits accruing to students in resource stressed locations, indicating that the improvement in mean achievement is independent of resource levels at a school. Only the results without controlling for prior

achievement are reported since the model that additionally included prior achievement – the test of marginal value-added achievement – less favorably supported Hypothesis 5 (all interactions terms had p values greater than 0.25) and offered less net improvement in the model fit (the quotient of the change in Deviance over the change in *df* was smaller).

Table 12a. Hypothesis 5 (School Resources): Three-Level HLM for Strong Test of the Dependence of the Effect of Years of Class Size Reduction Experience on Proxies for School Resource Disadvantages.

Predictors of Total Mathematics Achievement (SAT-9 Spring 1999)		Unstandardized Coefficient	Standard Error	p
Years of CSR Experience Effect		0.928	0.522	0.075
Predictors of Years of CSR Experience Effect				
School Proportion:	African American Students	-0.55	1.80	0.762
	Asian American Students	-1.35	7.47	0.857
	Hispanic Students	-2.53	2.50	0.313
	Low Income Students	0.09	1.95	0.965
	Full Credential Teachers	0.48	2.36	0.840
School Average:	Years Teaching Experience	0.0898	0.0639	0.160
		Δ Deviance	Δ *df*	p
	Model Change:	16.62	12	0.164

NOTE: Model change statistics (students within classrooms within schools) were calculated by comparing a base model that controls for implementation biases with a model that additionally enters Years of CSR experience, controlling for current CSR experience, and the six school-level predictors of the student-level Years of CSR effect; since the Deviance has a χ^2 distribution, a one-tail χ^2 test was used to obtain p.

Hence, if resources are being more effectively used in small class settings, the increase in effectiveness is virtually the same in high and low resource schools, making it impossible to assert that an overall improvement in classroom mean scores is in any way the result of teachers' greater capacity to utilize instructional resources effectively. Not only are there no statistically reliable coefficients among the resource stress indicators in Table 12, but three of them have the wrong sign to be considered as being supportive indicators. Two resource indicators (school proportion of African American students and school proportion of low income students) have

negative signs, though insignificant, while the near-significant (p = .108) coefficient for the proportion of fully credentialed teachers has a positive sign indicating that CSR just might be improving mean achievement in places with a more rather than less fully qualified staff. Quite simply, the data in this study offer no support for a resource effectiveness hypothesis.

Table 12b. Hypothesis 5 (School Resources): Three-Level HLM for Weak Test of the Dependence of Class Size Reduction Effect on Proxies for School Resource Disadvantages - Value Added by Class Size Reduction.

Predictors of Total Mathematics Achievement (SAT-9 Spring 1999)		Unstandardized Coefficient	Standard Error	*p*
Current CSR Experience Effect		3.71	0.82	0.000
Predictors of Current CSR Experience Effect				
School Proportion:	African American Students	4.83	6.65	0.467
	Asian American Students	15.84	16.35	0.333
	Hispanic Students	-0.68	7.23	0.925
	Low Income Students	-3.95	4.83	0.414
	Full Credential Teachers	-2.46	8.36	0.769
School Average:	Years Teaching Experience	0.250	0.204	0.222
		Δ Deviance	Δ *df*	*p*
	Model Change:	68.50	10	0.000

NOTE: Model change statistics (students within classrooms within schools) were calculated by comparing a base model that controls for implementation biases and prior achievement with a model that additionally enters current CSR experience and the six school-level predictors of the class-level CSR effect; since the Deviance has a χ^2 distribution, a one-tail χ^2 test was used to obtain *p*.

Finding #6: The Instructional Practices Hypothesis – there is small, but statistically reliable support for an inference that CSR improves instructional effectiveness.

Table 13 summarizes the HLM models testing for the impact of CSR on changing classroom-level patterns of student achievement. The four rows of this table are taken from four separate two-level HLM analyses, examining in succession whether small class experience contributes to changing classroom mean, standard deviations, skewness or kurtosis. In each case, intake patterns were also entered into the model in order to statistically equalize the class parameters before testing for CSR impact. Of course the mean remains significantly elevated by

CSR experience – this test is essentially a repetition of the mean achievement increase hypothesized and documented in findings 2, 3, 4 and 5 above. The only difference here is that the classroom intake patterns (prior mean, standard deviation, skewness and kurtosis) are included as control variables.

Table 13. Hypothesis 6 (Instructional Practices): Two-Level HLM for the Relationship of Class Size Reduction to the Classroom Outcome Patterns of Student Achievement on the Total Mathematics Battery for the Spring 1999 SAT-9.

Classroom Achievement Pattern Variable	Regression Analysis			Model Change Statistics		
	Unstandardized CSR Coefficient	Standard Error	*p*	Δ Deviance	Δ *df*	*p*
Mean	3.75	0.61	0.000	187.39	5	0.000
Standard Deviation[a]	-1.24	0.80	0.121	174.30	16	0.000
Skewness	-0.08	0.03	0.004	84.62	5	0.000
Kurtosis[b]	-0.14	0.05	0.008	14.35	5	0.000

NOTE: Model change statistics (classrooms within schools) were calculated by comparing a base model that controls for implementation biases with a model that additionally enters current CSR experience and the four classroom intake patterns of student achievement (i.e., prior [Spring 1998] achievement mean, standard deviation, skewness, and kurtosis); since the Deviance has a χ^2 distribution, a one-tail χ^2 test was used to obtain *p*.

[a] Classroom standard deviation is the only pattern variable with significant (p<.05) CSR interaction terms; there are two: the unstandardized CSR by classroom proportion DIS coefficient is 12.81 (std. err. 6.23), and the unstandardized CSR by classroom proportion African American coefficient is 2.70 (std. err. 1.32).

[b] Because classroom kurtosis has no significant school level variance component, the statistics reported here were calculated using standard regression analysis at the classroom level; in this case the model change was evaluated using the change in F (df_2=670) rather than the Deviance statistic.

The second row in Table 13 indicates that CSR had no significant impact on class standard deviations, confirming that classrooms with CSR experience have about the same amount of dispersion around the mean as the large classes. As indicated in the third and fourth rows of the table, however, CSR experience is associated with reliable (though small) changes in class skewness and kurtosis. Conceptually, these statistical findings indicate that teachers in small classes were able to shift the performance of the bulk of their middle-performing students

toward the performance of the best students in the class. The lowest performing students were helped less than the middle-performers, increasing the negative skew. Since lowering the kurtosis means reducing the number of outliers, it is appropriate to conclude that the highest performers served as "attractors" or role-models for the mid-range students, reducing the probability that classes would have high performing outliers. Since this change in pattern was accompanied by an overall increase in mean achievement, it is fair to assume that, with CSR implementation, we are seeing a slight shift in who benefits most from effective teaching toward the middle-performing students.

Conclusion

This study has tested six core hypotheses regarding the impact of California' Class Size Reduction on student academic achievement as measured by mandated Stanford Achievement Test, 9th edition (SAT-9). The first hypothesis – that the programmatic character of California's CSR initiative led to significant implementation biases, providing different CSR reduction patterns to very different groups of children – proved to be the most robust. More than 54 percent of the variations in CSR exposure can be explained by nineteen variables reflecting student demographic characteristics, classroom assignments and teacher characteristics. The second hypothesis – that CSR significantly impacts student achievement – was proven relatively weak. Achievement impacts in reading and language sub-tests were virtually non-existent. Those for mathematics while substantial for some types of CSR experience were quite varied and inconsistent in overall effect.

Hypothesis 3, that smaller classes facilitate more effective socialization of children to the school culture, especially during their first critical years of in school, was supported to the extent that CSR appeared to make greater contributions to mathematics achievement for children who

started in first grade. The corollary hypothesis that early grade CSR should be most effective in raising the achievement of children facing the greatest socialization challenges was not supported in any of eight tests. Early CSR raised math achievement generally, but had no special impact on ethnic minority groups, children from Spanish or other non-English speaking homes or children who qualify for free or reduced price lunch.

Hypothesis 4, suggesting that CSR may be effective because it makes it easier for teachers to manage instruction for harder to teach children was also supported in only the most general way. Mathematics scores were significantly higher for children who were in small classes during the year they were tested, but CSR provided no special advantages to students who found themselves in classrooms with more than two difficult to teach children. The expectation that teachers would be better able to cope with large numbers of special education children, or with larger numbers of non-English speakers or any of the other challenging conditions tested in this model found no significant benefits accruing to smaller class participants.

Hypothesis 5, exploring whether smaller classes might be more helpful in resource poor environments was thoroughly disappointed. Smaller class size made no special contribution to the achievement of children facing any of the resource limitations associated with being in schools impacted by poverty, ethnic group concentrations, or inexperienced teachers.

Hypothesis 6, suggesting that CSR might be affecting achievement by enabling teachers to change the pattern of attainment among the group of students assigned to them did receive some statistically reliable support in that smaller classes had not only higher mean scores, but also more negative skewness in the distribution of scores and a bit of reduction in the number of outlier students. That is, small class teachers did, on the average, produce mathematics

achievement profiles for their students that moved the bulk of the middle range students closer to the highest performing individual students in their classes. Here again, however, shifts in the pattern of achievement were so modest as to raise significant questions about whether these small changes could possibly justify the enormous amounts of money being poured into shrinking class sizes.

We conclude by reiterating the cautionary note that, within our study sample, CSR is seriously confounded with student grade cohort. Only fourth graders in our sample had no CSR experience, and no third graders had only one year of CSR. On the mathematics test, the third grade outperformed the fourth grade, and it is not possible to be certain whether this was the result of their much higher rate of participation in CSR or because there is simply a year-to-year cohort difference in the average attainment level of the two student cohorts. We did not statistically equalize the third and fourth graders in the analysis presented in this report because we suspect that the third graders may be outperforming the fourth graders just because of their greater exposure to CSR. When we did statistically remove the third grade advantage, the apparent CSR effect was reduced by more than two-thirds. Continued study of these students through at least one more academic year will enable us to isolate the grade cohort effect and reliably separate it from the CSR effect.

REFERENCES

Achilles, C. M. (1999). *Let's put kids first, finally: Getting class size right.* Thousand Oaks, CA: Corwin Press.

Alexander, K. L., Entwisle, D. R., & Thompson, M. S. (1987). School performance, status relations, and the structure of sentiment: Bringing the teacher back in. *American Sociological Review, 52*(5), 665-682.

Anderson, L. W. (1999). *Why should reduced class size lead to increased student achievement?* Paper presented at the National Invitational Conference on How Small Classes Help Teachers Do Their Best, Washington, DC.

Arum, R. (2000). Schools and communities: Ecological and institutional dimensions. *Annual Review of Sociology, 26*, 395-418.

Bain, H. P., Lintz, N., & Word, E. (1989). A study of first grade effective teaching practices from the Project STAR class-size research. *Resources in Education.* (ERIC Document Reproduction Service No. ED321887.)

Balow, I. H., & Schwager, M. (1990). *Retention in grade: A failed procedure.* Riverside, CA: University of California, Riverside, California Educational Research Cooperative.

Black, S. E. (1999). Do better schools matter? Parental valuation of elementary education. *The Quarterly Journal of Economics, 114*(2), 577-599.

Blatchford, P. (2000). Summary of symposium to be presented at the British Educational Research Association Annual Conference, Cardiff, 2000. Personal communication received September 28, 2000.

Blatchford, P., Goldstein, H., & Mortimore, P. (1998). Research on class size effects: A critique of methods and a way forward. *International Journal of Educational Research, 29*(8), 691-710.

Bohrnstedt, G. W., & Stecher, B. M. (Eds.) (1999). Class size reduction in California: Early evaluation findings, 1996-1998. [*CSR research consortium, year 1 evaluation report*] Palo Alto, CA: American Institutes for Research.

Bryk, A. S. & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods.* Newbury Park, CA: Sage Publications.

Bryk, A. S., Raudenbush, S. W., & Congdon, R. T. (2000). *HLM5: Hierarchical linear and nonlinear modeling*. Chicago: Scientific Software International.

Burns, R. (1996). *A study of combination class achievement.* Riverside, CA: University of California, Riverside, California Educational Research Cooperative.

Burns, R. B., & Mason, D. A. (1998). Class formation and composition in elementary schools. *American Educational Research Journal, 35*(4), 739-772.

Cahen, L. S., Filby, N., McCutcheon, G., & Kyle, D. W. (1983). *Class size and instruction.* New York: Longman.

Clotfelter, C. T. (1998). Public school segregation in metropolitan areas [*NBER Working Paper 6779*]. Cambridge, MA: National Bureau of Economic Research.

Darling-Hammond, L. (1998). Teachers and teaching: Testing policy hypotheses from a national commission report. *Educational Researcher, 27*(1), 5-15.

Entwisle, D. R. & Alexander, K. L. (1992). Summer setback: race, poverty, school composition and mathematics achievement. *American Sociological Review, 57*(1), 72-84.

Entwisle, D. R., Alexander, K. L., & Olson, L. S. (1997). *Children, schools, and inequality.* Boulder, CO: Westview Press.

Finn, J. D. (1998). *Class size and students at risk: What is known? What is next?* Paper commissioned by the National Institute on the Education of At-Risk Students, Office of Educational Research and Improvement, U.S. Department of Education, Washington, DC.

Finn, J. D., & Achilles, C. M. (1990). Answers and questions about class size: A statewide experiment. *American Educational Research Journal, 27*(3), 557-577.

———. (1999). Tennessee's class size study: Findings, implications, misconceptions. *Educational Evaluation and Policy Analysis, 21*(2), 97-109.

Finn, J. D., Gerber, S. B., Achilles, C. M., & Boyd-Zaharias, J. (in press). The enduring effects of small classes. *Teachers College Record.*

Folger, J., & Breda, C. (1989). Evidence from Project STAR about class size and student achievement. *Peabody Journal of Education, 67*(1), 17-33 (published in 1992).

Glass, G. V., Cahen, L. S., Smith, M. L., & Filby, N. N. (1982). *School class size: Research and policy.* Beverly Hills, CA: Sage Publications.

Goldstein, H. (1995). *Multilevel statistical models, 2nd edition.* London: Edward Arnold.

Goldstein, H. & Blatchford, P. (1997). *Class size and educational achievement: A methodological review.* Report commissioned by UNESCO, Geneva.

Hakuta, K., Butler, Y. G., & Witt, D. (January, 2000). How long does it take English learners to attain proficiency? [*Policy Report 2000-1*]. Santa Barbara, CA: The University of California Language Minority Research Institute.

Hallinan, M. T., & Sorensen, A. B. (1985). Class size, ability group size, and student achievement. *American Journal of Education, 94*(1), 71-89.

Han, L. & Hoover, H. D. (1994). *Gender differences in achievement test scores.* Paper presented at the Annual Meeting of the National Council on Measurement in Education, New Orleans, Louisiana.

Hanushek, E. A. (1998). The evidence on class size. [*Occasional Paper Number 98-1*]. Rochester, NY: University of Rochester, W. Allen Wallis Institute of Political Economy.

———. (1999). Some findings from an independent investigation of the Tennessee STAR experiment and from other investigations of class size effects. *Educational Evaluation and Policy Analysis, 21*(2), 143-163.

Hargeaves, L, Galton, M., & Pell, A. (1997). *The effects of major changes in class size on teacher-pupil interaction in elementary school classes in England. Does research merely confirm the obvious?* Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, Illinois.

Hruz, T. (2000). The costs and benefits of smaller classes in Wisconsin: A further evaluation of the SAGE program. *Wisconsin Policy Research Institute Report, 13*(6), 1-44.

Hymon, S. (1997). A lesson in classroom size reduction. *School Planning and Management, 36*(7), 18-23.

Illig, D. C. (1997). *Early implementation of the class size reduction initiative.* Sacramento, CA: California Research Bureau, California State Library.

Jencks, C. & Phillips, M. (Eds.). (1998). *The black-white test score gap.* Washington, DC: Brookings Institution Press.

Johnston, J. M. (1989). Teacher perceptions of changes in teaching when they have a small class or an aide. *Peabody Journal of Education, 67*(1), 106-122 (published in 1992).

Krueger, A. B. (1999). Experimental estimates of education production functions. *The Quarterly Journal of Economics, 114*(2), 497-532.

Krueger, A. B. & Whitmore, D. M. (2001). The effect of attending a small class in the early grades on college-test taking and middle school test results: Evidence from Project STAR. *The Economic Journal, 111*(468), 1-28.

Mitchell, D. E., Beach, S. A., & Badarak, G. (1989) Modeling the relationship between achievement and class size: A re-analysis of the Tennessee Project STAR data. *Peabody Journal of Education, 67*(1), 34-74 (published in 1992).

Mitchell, D. E., Destino, T., and Karam, R. (1997). *Evaluation of English language development programs in the Santa Ana Unified School District: A report on data system reliability and statistical modeling of program impacts.* Riverside, CA: University of California, Riverside, California Educational Research Cooperative.

Mitchell, D. E., & Mitchell, R. E. (1999). *The impact of California's class size reduction initiative on student achievement: Detailed finding from eight school districts.* Riverside, CA: University of California, Riverside, California Educational Research Cooperative.

———. (2001). *The impact of California's class size reduction initiative on student achievement: Third year evaluation report.* Riverside, CA: University of California, Riverside, California Educational Research Cooperative.

Mitchell, R. E. (2000). *The effects of California's Class Size Reduction Program on elementary school mathematics achievement.* Unpublished manuscript, California Educational Research Cooperative, University of California, Riverside, California.

Mitchell, R. E., & Mitchell, D. E. (1999). *Student segregation and achievement tracking in year-round schools.* Paper presented at the Annual Meeting of the American Sociological Association, Chicago, Illinois.

Molnar, A., Smith, P., & Zahorik, J. (with K. Ehrle, A. Halbach, A. Palmer, & B. Schoeller) (1999). *1998-99 Evaluation Results of the Student Achievement Guarantee in Education (SAGE) Program.* Milwaukee, WI: University of Wisconsin, Milwaukee, School of Education.

Nye, B., Hedges, L. V., & Konstantopoulos, S. (1999). The long-term effects of small classes: A five-year follow-up of the Tennessee class size experiment. *Educational Evaluation and Policy Analysis, 21*(2), 127-142.

———. (2000). The effects of small classes on academic achievement: The results of the Tennessee class size experiment. *American Educational Research Journal, 37*(1), 123-151.

Ogawa, R. T., Huston, D., & Stine, D. E. (1999). California's class size reduction initiative: Differences in teacher experience and qualifications across districts. *Educational Policy, 13*(5), 659-673.

Pate-Bain, H., Boyd-Zaharias, J., Cain, V. A., Word, E., & Binkley, M. E. (1997). *STAR follow-up studies 1996-97.* Lebanon, TN: HEROS, Inc., 1997.

Rosenthal, A. S., Baker, K., & Ginsburg, A. (1983). The effect of language background on achievement level and learning among elementary school students. *Sociology of Education, 56*(4), 157-169.

Rowan, B. & Miracle, Jr., A. W. (1983). Systems of ability grouping and the stratification of achievement in elementary schools. *Sociology of Education, 56*(3), 133-144.

Shields, C. M., & Oberg, S. L. (2000). *Year-round schooling: Promises & pitfalls.* Lanham, MD: The Scarecrow Press.

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling.* Thousand Oaks, CA: Sage Publications.

SPSS Inc. (1999). *SPSS for Windows, Version 9.0.* Chicago, IL: SPSS Inc. [contractor/manufacturer]

Stecher, B. M., & Bohrnstedt, G. W. (Eds.). (2000). *Class size reduction in California: The 1998-99 evaluation findings.* Sacramento, CA: California Department of Education.

Veenman, S. (1995). Cognitive and noncognitive effects of multigrade and multi-age classes: A best-evidence synthesis. *Review of Educational Research, 65*(4), 319-381.

Wexler, E., Izu, J., Carlos, L., Fuller, B., Hayward, G., & Kirst, M. (1998). *California's class size reduction: Implications for equity, practice & implementation.* San Francisco, CA: PACE-WestEd.

Wright, S. P., Horn, S. P., & Sanders, W. L. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education, 11*, 57-67.

Zahorik, J. A. (1999). Reducing class size leads to individualized instruction. *Educational Leadership, 57*(1), 50-53.

APPENDIX A

Variables Analyzed in this Study

Dependent Variables. The dependent variables – reading, mathematics and language achievement – were measured using 1999 Normal Curve Equivalent (NCE) scores from the 9th Edition (Form T) of the Stanford Achievement Test (SAT-9) as mandated by the California Department of Education. In addition to assessing the specific impact of California's Class Size Reduction (CSR) initiative, this report examines the effect of student background, classroom context and teacher characteristics on individual achievement levels (i.e., Total Reading, Total Mathematics and Total Language SAT-9 scores).

Independent Variables. The central independent variable of interest in this study is, of course, class size – the number of students assigned to each teacher. We seek to determine the extent to which providing children in kindergarten through grade three with classes that have a maximum of 20 students (rather than the 28 to 32 students typical of California public schools prior to the adoption of CSR) has a positive impact on their learning. Class size is not the only influence on student learning, however. Painstaking, and often quite expensive, efforts to improve public school performance over the past several decades has taught us that student achievement is shaped by a broad range of potent demographic, social and schooling factors – factors that are often very unevenly distributed across classrooms, schools or school districts.

In the study reported here, 20 covariates with potentially powerful impacts on student academic achievement are examined. Sixteen additional variables defining classroom environmental contexts were generated by calculating classroom proportions for each factor level of seven demographic and classroom assignment variables. Taken together, these 36 variables surround and embed student achievement in five distinct contexts or three hierarchical levels. The five contexts are depicted in Figure I. At the first level (1A) – Student Demography – five factors, with dummy-coded levels, constitute the most fundamental and intractable academic performance influences: gender (two levels: female=1, male=0), family poverty (three levels: not qualified for National School Lunch Program reference category for free lunch qualifed=1, and reduced price lunch qualified=1), ethnicity (five levels: White the reference category for African American=1, Asian American=1, Hispanic=1, and Other Non-White=1), home language (three levels: English the reference category for Spanish=1, and Other Non-English=1), and time of admission to the local school (two levels: New to School in 1998-99=1, Not New=0).

At level 1B, school organizations begin their influence on student academic opportunities by making class assignments. Five factors – grade level assignment (Grade 3=1, Grade 4=0), grade retention resulting in overage students (two levels: Overage 15+ Months=1, Not Overage=0), English language proficiency assessment (three levels: English Only the reference category for Limited English Proficient [LEP]=1, and Fluent English Proficient [FEP]=1), special education certification (four levels: Not Identified for Special Education reference category for Resource Specialist Program [RSP]=1, Designated Instructional Services [DIS]=1, and Gifted and Talented Education [GATE]=1), and the level of placement in combination grade classes (three levels: Not in Combo Class reference category for Low Grade in Combo Class=1, and High Grade in Combo Class=1) – are the most obvious ***classroom assignment indicators***.

Classroom environments constitute the third context level, Level 2A. Classroom environments are very complex and difficult to assess precisely. They are represented in this study by several variables. Two variables of our study operate only at the classroom level – year

round education track assignment and whether schools utilize combination grade classes. Additionally, this study examines fifteen calculated "concentration variables" that help to define the classroom environment by measuring the classroom proportions of:

Gender:
1. a single gender (girls),

Family income status:
2. low income status or "poverty" (children on the National School Lunch Program),

Retention in grade proxy:
3. overage-for-grade students (15+ months above a September start date for their grade),

Ethnic groups:
4. African-American (black) students
5. Hispanic students
6. Asian students
7. Other non-White students

Different home language groups:
8. Spanish home language speakers
9. Other non-English home language speakers

English language fluency groups:
10. Fluent English Proficient (FEP) students
11. Limited English Proficient (LEP) students

Special education category groups:
12. Resource Specialist Program (RSP – educationally at risk) students
13. Designated Instructional Service (DIS – blind, deaf, speech impaired, physically handicapped, etc.) students, and
14. Gifted and Talented Education (GATE) students

Intra-district transiency
15. Proportion of students in the classroom that are new to the school in the test year

Teacher characteristics comprise the fourth context of influence over student achievement, Level 2B. Interacting with and potentially confounding the impact of class size, we would expect to find significant influence from teacher credentials (two level: Not Fully Credentialed=1, Fully Credentialed=0), education levels (three levels: BA + 30 or semester hours reference category for BA with less than 30 additional semester hours=1, and MA or Higher=1), and years of experience as well as from teacher gender (male=1, female=0), ethnicity (four levels: White reference category for African American=1, Hispanic=1, and Other Non-White=1), age in years and contract status (four levels: Tenure contract reference category for Probationary=1, Long-Term Substitute/Temporary=1, and Other Contract=1).

After these variables are all controlled (using statistical procedures to remove their impact on achievement because experimental controls are not available), we would still expect unmeasured **school level factors** to have some influence on student achievement. At this level, we only need to examine the extent to which the unmeasured influences associated with student attendance boundaries remain powerful, and to statistically remove them. School level aggregates of all of the aforementioned variables are available to specific hypotheses that require their specification.

APPENDIX B
Details of HLM Models

The dependent variable in the HLM analyses is the standardized total subject achievement for mathematics on the SAT-9 scaled in Normal Curve Equivalents (NCE scores). This provides a common standardized metric for third and fourth grade students – performance relative to the national norms. For three-level models, individual student achievement scores are the dependent variable. For the two-level model of classroom outcomes (Hypothesis 6), the four separate models have the classroom mean, standard deviation, skewness, and kurtosis, respectively, as the dependent variable. For generic simplicity, the dependent variable will be referred to as achievement and labeled "*Ach*" (subscripted i, j, and k for student-, classroom-, and school-level, respectively) or "*ClassAchMoment*" (subscripted i and j for classroom- and school-levels, respectively).

As described in Appendix A, there are a number of student- and classroom-level covariates whose presence is required to control for implementation biases. These will be referred to as the vector of covariates at each level. The vector will be summarized as a sum and written as $\Sigma_l(\pi_{ljk}a_{ijk})$, with l being the subscript for each of the L covariates at the student-level and $\Sigma_m(\beta_{mk}x_{jk})$, with m being the subscript for each of the M covariates at the classroom-level.

Each π (student-level) coefficient and β (class-level) coefficient that has no higher level predictors or random effect terms is equal to its corresponding school-level fixed effect (γ) and separate equations will not be written out below.

Hypothesis 2:

$$\begin{aligned}
Ach_{ijk} = {} & \pi_{0jk} + \Sigma_l(\pi_{ljk}a_{ijk}) + \pi_{YearsCSRExperience,jk}YearsCSRExperience_{ijk} + \\
& \pi_{YearsCSRExperience x AfroAmer,jk}YearsCSRExperience_{ijk}xAfroAmer_{ijk} + \\
& \pi_{YearsCSRExperience x AsianAmer,jk}YearsCSRExperience_{ijk}xAsianAmer_{ijk} + \\
& \pi_{YearsCSRExperience x Hispanic,jk}YearsCSRExperience_{ijk}xHispanic_{ijk} + \\
& \pi_{YearsCSRExperience x OtherEthn,jk}YearsCSRExperience_{ijk}xOtherEthn_{ijk} + \\
& \pi_{YearsCSRExperience x NSLP(free),jk}YearsCSRExperience_{ijk}xNSLP(free)_{ijk} + \\
& \pi_{YearsCSRExperience x NSLP(reduced),jk}YearsCSRExperience_{ijk}xNSLP(reduced)_{ijk} + \\
& \pi_{YearsCSRExperience x Spanish,jk}YearsCSRExperience_{ijk}xSpanish_{ijk} + \\
& \pi_{YearsCSRExperience x OtherLang,jk}YearsCSRExperience_{ijk}xOtherLang_{ijk} + e_{ijk}
\end{aligned}$$

$$\pi_{0jk} = \beta_{00k} + \Sigma_m(\beta_{mk}x_{jk}) + \beta_{0,CurrentCSR,k}CurrentCSR_{jk} + r_{0jk}$$

$$\beta_{00k} = \gamma_{000} + u_{00k}$$

$$\beta_{0,CurrentCSR,k} = \gamma_{0,CurrentCSR,0} + u_{0,CurrentCSR,k}$$

Hypothesis 3:

$$\begin{aligned}
Ach_{ijk} = {} & \pi_{0jk} + \Sigma_l(\pi_{ljk}a_{ijk}) + \pi_{Start1stGrade,jk}Start1stGrade_{ijk} + \pi_{Start2ndGrade,jk}Start2ndGrade_{ijk} + \\
& \pi_{Start1stGrade x AfroAmer,jk}Start1stGrade_{ijk}xAfroAmer_{ijk} + \\
& \pi_{Start1stGrade x AsianAmer,jk}Start1stGrade_{ijk}xAsianAmer_{ijk} + \\
& \pi_{Start1stGrade x Hispanic,jk}Start1stGrade_{ijk}xHispanic_{ijk} + \\
& \pi_{Start1stGrade x OtherEthn,jk}Start1stGrade_{ijk}xOtherEthn_{ijk} + \\
& \pi_{Start1stGrade x NSLP(free),jk}Start1stGrade_{ijk}xNSLP(free)_{ijk} + \\
& \pi_{Start1stGrade x NSLP(reduced),jk}Start1stGrade_{ijk}xNSLP(reduced)_{ijk} + \\
& \pi_{Start1stGrade x Spanish,jk}Start1stGrade_{ijk}xSpanish_{ijk} +
\end{aligned}$$

$\pi_{Start1stGrade x OtherLang,jk}Start1stGrade_{ijk}xOtherLang_{ijk}$
$\pi_{Start2ndGrade x AfroAmer,jk}Start2ndGrade_{ijk}xAfroAmer_{ijk}$ +
$\pi_{Start2ndGrade x AsianAmer,jk}Start2ndGrade_{ijk}xAsianAmer_{ijk}$ +
$\pi_{Start2ndGrade x Hispanic,jk}Start2ndGrade_{ijk}xHispanic_{ijk}$ +
$\pi_{Start2ndGrade x OtherEthn,jk}Start2ndGrade_{ijk}xOtherEthn_{ijk}$ +
$\pi_{Start2ndGrade x NSLP(free),jk}Start2ndGrade_{ijk}xNSLP(free)_{ijk}$ +
$\pi_{Start2ndGrade x NSLP(reduced),jk}Start2ndGrade_{ijk}xNSLP(reduced)_{ijk}$ +
$\pi_{Start2ndGrade x Spanish,jk}Start2ndGrade_{ijk}xSpanish_{ijk}$ +
$\pi_{Start2ndGrade x OtherLang,jk}Start2ndGrade_{ijk}xOtherLang_{ijk} + e_{ijk}$
$\pi_{0jk} = \beta_{00k} + \Sigma_m(\beta_{mk}x_{jk}) + r_{0jk}$
$\beta_{00k} = \gamma_{000} + u_{00k}$

Hypothesis 4:
$Ach_{ijk} = \pi_{0jk} + \Sigma_l(\pi_{ljk}a_{ijk}) + \pi_{PriorAch,jk}PriorAch_{ijk} + e_{ijk}$
$\pi_{0jk} = \beta_{00k} + \Sigma_m(\beta_{mk}x_{jk}) + \beta_{0,CurrentCSR,k}CurrentCSR_{jk}$ +
$\beta_{0,CurrentCSR x ClassPropDIS,k}CurrentCSR_{jk}xClassPropDIS_{jk}$ +
$\beta_{0,CurrentCSR x ClassPropRSP,k}CurrentCSR_{jk}xClassPropRSP_{jk}$ +
$\beta_{0,CurrentCSR x ClassPropAfroAmer,k}CurrentCSR_{jk}xClassPropAfroAmer_{jk}$ +
$\beta_{0,CurrentCSR x ClassPropAsianAmer,k}CurrentCSR_{jk}xClassPropAsianAmer_{jk}$ +
$\beta_{0,CurrentCSR x ClassPropHispanic,k}CurrentCSR_{jk}xClassPropHispanic_{jk}$ +
$\beta_{0,CurrentCSR x ClassPropOtherEthn,k}CurrentCSR_{jk}xClassPropOtherEthn_{jk}$ +
$\beta_{0,CurrentCSR x ClassPropSpanish,k}CurrentCSR_{jk}xClassPropSpanish_{jk}$ +
$\beta_{0,CurrentCSR x ClassPropOtherLang,k}CurrentCSR_{jk}xClassPropOtherLang_{jk}$ +
$\beta_{0,CurrentCSR x ClassPropNSLP,k}CurrentCSR_{jk}xClassPropNSLP_{jk} + r_{0jk}$
$\pi_{PriorAch,jk} = \beta_{PriorAch,0k} + r_{PriorAch,jk}$
$\beta_{00k} = \gamma_{000} + u_{00k}$
$\beta_{0,CurrentCSR,k} = \gamma_{0,CurrentCSR,0} + u_{0,CurrentCSR,k}$
$\beta_{PriorAch,0k} = \gamma_{PriorAch,00} + u_{PriorAch,0k}$

Hypothesis 5:
"Strong Test"
$Ach_{ijk} = \pi_{0jk} + \Sigma_l(\pi_{ljk}a_{ijk}) + e_{ijk}$
$\pi_{0jk} = \beta_{00k} + \Sigma_m(\beta_{mk}x_{jk}) + \beta_{0,CurrentCSR,k} + r_{0jk}$
$\beta_{00k} = \gamma_{000} + u_{00k}$
$\beta_{0,CurrentCSR,k} = \gamma_{0,CurrentCSR,0} + \gamma_{0,CurrentCSR,SchPropAfroAmer}SchPropAfroAmer_k$ +
$\gamma_{0,CurrentCSR,SchPropAsianAmer}SchPropAsianAmer_k$ +
$\gamma_{0,CurrentCSR,SchPropHispanic}SchPropHispanic_k$ +
$\gamma_{0,CurrentCSR,SchPropNSLP}SchPropNSLP_k$ +
$\gamma_{0,CurrentCSR,SchPropFullCredTchrs}SchPropFullCredTchrs_k$ +
$\gamma_{0,CurrentCSR,SchAvgYrsTchrExp}SchAvgYrsTchrExp_k + u_{0,CurrentCSR,k}$

"Weak Test"
$Ach_{ijk} = \pi_{0jk} + \Sigma_l(\pi_{ljk}a_{ijk}) + e_{ijk}$
$\pi_{0jk} = \beta_{00k} + \Sigma_m(\beta_{mk}x_{jk}) + \beta_{0,CurrentCSR,k} + r_{0jk}$
$\beta_{00k} = \gamma_{000} + u_{00k}$

$$\beta_{0,CurrentCSR,k} = \gamma_{0,CurrentCSR,0} + \gamma_{0,CurrentCSR,SchPropAfroAmer}SchPropAfroAmer_k + \gamma_{0,CurrentCSR,SchPropAsianAmer}SchPropAsianAmer_k + \gamma_{0,CurrentCSR,SchPropHispanic}SchPropHispanic_k + \gamma_{0,CurrentCSR,SchPropNSLP}SchPropNSLP_k + \gamma_{0,CurrentCSR,SchPropFullCredTchrs}SchPropFullCredTchrs_k + \gamma_{0,CurrentCSR,SchAvgYrsTchrExp}SchAvgYrsTchrExp_k + u_{0,CurrentCSR,k}$$

Hypothesis 6:

$$\mathit{ClassAchMoment}_{jk} = \beta_{0k} + \Sigma_m(\beta_{mk}x_{jk}) + \beta_{CurrentCSR,k}CurrentCSR_{jk} + \beta_{PriorClassMeanAch,k}PriorClassMeanAch_{jk} + \beta_{PriorClassStdDev,k}PriorClassStdDev_{jk} + \beta_{PriorClassSkewness,k}PriorClassSkewness_{jk} + \beta_{PriorClassKurtosis,k}PriorClassKurtosis_{jk} + \beta_{CurrentCSRxClassPropDIS,k}CurrentCSR_{jk}xClassPropDIS_{jk} + \beta_{CurrentCSRxClassPropRSP,k}CurrentCSR_{jk}xClassPropRSP_{jk} + \beta_{CurrentCSRxClassPropAfroAmer,k}CurrentCSR_{jk}xClassPropAfroAmer_{jk} + \beta_{CurrentCSRxClassPropAsianAmer,k}CurrentCSR_{jk}xClassPropAsianAmer_{jk} + \beta_{CurrentCSRxClassPropHispanic,k}CurrentCSR_{jk}xClassPropHispanic_{jk} + \beta_{CurrentCSRxClassPropOtherEthn,k}CurrentCSR_{jk}xClassPropOtherEthn_{jk} + \beta_{CurrentCSRxClassPropSpanish,k}CurrentCSR_{jk}xClassPropSpanish_{jk} + \beta_{CurrentCSRxClassPropOtherLang,k}CurrentCSR_{jk}xClassPropOtherLang_{jk} + \beta_{CurrentCSRxClassPropNSLP,k}CurrentCSR_{jk}xClassPropNSLP_{jk} + r_{jk}$$

$$\beta_{0k} = \gamma_{00} + u_{0k}$$

where *Moment* is the first (mean), second (standard deviation), third (skewness), or fourth (kurtosis) moment of the achievement distribution – the class pattern variables.

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC®

Reproduction Release

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: Competing Explanations of Class Size Reduction Effects: The California Case.	
Author(s): Douglas E. Mitchell and Ross E. Mitchell	
Corporate Source: California Educational Research Cooperative	Publication Date: April 14, 2001

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, Resources in Education (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign in the indicated space following.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY SAMPLE TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY SAMPLE TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY SAMPLE TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
Level 1	Level 2A	Level 2B
X		
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g. electronic) *and* paper copy.	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche, or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Signature: Ross E Mitchell	Printed Name/Position/Title: Ross E. Mitchell, Research Scientist	
Organization/Address: Gallaudet Research Institute, HMB S-428 Gallaudet University 800 Florida Ave, NE Washington, DC 20002-3695	Telephone: 202-651-5576	Fax: 202-651-5746
	E-mail Address: ross.mitchell@gallaudet.edu	Date: February 4, 2002

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, or, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

ERIC Clearinghouse on Educational Management
1787 Agate Street
5207 University of Oregon
Eugene, OR, 97403-5207
attn: Acquisitions